DOCUMENT RESUME

ED 115 862 CE 005 700

AUTHOR	Creager, John A.
TITLE	The Data Entry System for the Sample Data System of MISOE [Management Information System for Occupational Education]: Operations Report No. 1.
INSTITUTION	Management and Information System for Occupational Education, Winchester, Mass.
SPONS AGENCY	Massachusetts State Dept. of Education, Boston. Div. of Occupational Education.
NOTE	95p.; For related documents, see ED 062 553; ED 068 646-647; ED 072 225; ED 072 228; ED 072 303-304; CE 005 687-727; This document is Attachment 7 referred to in CE 005 699
EDRS PRICE	MF-$0.76 HC-$4.43 Plus Postage
DESCRIPTORS	Computer Programs; *Computer Storage Devices; Data Processing; Demonstration Projects; *Electronic Data Processing; *Input Output Devices; Instrumentation; *Management Information Systems; Optical Scanners; State Programs; *Vocational Education
IDENTIFIERS	Management Information System Occupational Educa; MISOE; Sample Data System; SDS

ABSTRACT

The report specifies the design of the data entry system for the Sample Data System subsystems (SDS-1 and SDS-2) providing connectability to meet previously specified purposes and functions of the Management Information System for Occupational Education (MISOE). General assumptions for specifying the SDS data entry system are stated including the important assumption that initial processing of information be done by optical scanning using an OPSCAN-100 (Digitek) and 24K Honeywell computer system. General specifications for operational MISOE files interfacing the analysis system are presented, with the file identification-descriptor system designed to ensure connectability of MISOE components. The specifications for the individual file types as they are developed longitudinally are also given. The general considerations for moving from available instrumentation to operational MISOE SDS files are presented using flow charts and accompanying text to provide an overview of the SDS data entry system. More precise specifications of optical scanning and of the data entry operations for the individual pieces of data entering the system from the following batteries are provided: student input, student process, student product, impact, teacher, and administrator batteries. (Author/MS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED115862

OPERATIONS REPORT NO. 1

The Data Entry System for the Sample Data System of MISOE

John A. Creager

Management Information System for Occupational Education
Division of Occupational Education
1017 Main Street
Winchester, Massachusetts 01890

CE005700

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Table of Contents

Page

I. Introduction... 1

II. SDS Files and Their Identification-Descriptor System........... 5
- A. Introduction... 5
- B. The Identification-Descriptor System...................... 6
- C. The Confidentiality System................................ 8
- D. The Name and Address Files................................ 10
- E. The "Program" File.. 13
- F. The Student Data File..................................... 14
- G. The Teacher Data File..................................... 17
- H. The Administrator Data File............................... 19
- I. The Special Cross-Sectional Impact Files.................. 20

III. From Instruments to MISOE Master Files...................... 22
- A. Introduction... 22
- B. The Data Collection Phase................................. 22
- C. The Optical Scanning Phase................................ 28
- D. Operations on Input Data Tapes............................ 32
- E. Battery Merge and ID Change: Getting to the Master File... 37
- F. Debugging Operations...................................... 37

IV. Data Entry Operations With The Student Input Battery........... 38
- A. The SMIF.. 39
- B. The ITED File... 42
- C. The DAT File.. 45
- D. The SSHA File... 46
- E. The Personal Values File.................................. 47
- F. The HSPQ File... 49
- G. The Culture Fair File..................................... 50
- H. The MPI File.. 50

V. Data Entry Operations With the Student Process Battery......... 60
- A. The School Sentiment Index Files.......................... 62
- B. The Student Program Questionnaire Files................... 65
- C. The Master Identification Form Update Files............... 66

VI. Data Entry Specifications for the Student Product Battery...... 66

VII. Data Entry Specifications for the Impact Batteries............ 73
- A. Longitudinal Logistics for the MEII....................... 74
- B. Coding and Editing Specifications for the MEII............ 76
- C. Longitudinal Logistics for the MJEF....................... 84
- D. Battery Merge Problems for MEII and MJEF: Respondents and Nonrespondents....................... 86
- E. A Final Suggestion for Reducing the Load in Impact Space... 86

VIII. Data Entry Specifications for the Teacher and Administrator Batteries.. 87
- A. Specifications for Processing Answer Sheets for the Cover Sheet, TMIF, and AMIF.......................... 88
- B. The Planning Activities Sheet............................. 88
- C. The Image of Vocational Education......................... 88
- D. The Teacher Program Questionnaire......................... 89
- E. The Purdue Teacher Opinionaire............................ 89
- F. The MOETS and MAI... 90

List of Figures and Tables

Page

Figure 1: Schematic Layout of Identification-Descriptor Section of MISOE Tapes............ 6

Figure 2: Schematic Layout of the Name and Address Files................................ 12

Figure 3: Schematic Layout of Student Data Master Files 15

Table 1: Coding Specifications for the Items in the Massachusetts Pupil Inventory................ 52

Table 2: Coding Specifications for the Items in the Massachusetts Adult Pupil Inventory.......... 58

Table 3: Coding Specifications for the School Sentiment Index.............................. 64

Table 4: Coding Specifications for the MEII........... 79

Table 5: Data Entry Specifications for the MJEF....... 85

Table 6: Data Entry Specifications for MOETS and MAI.. 91

The Data Entry System for the Sample Data System of MISOE

I. Introduction

2

The design of the Management Information System for Occupational Education (MISOE) includes designs of subsystems and their connectabilities to meet previously specified purposes and functions. Such sybsystems include the data systems: the Census Data System (CDS), and the two Sample Data Systems (SDS-1 and SDS-2); the analysis systems: static and dynamic; the data entry system; and the information retrieval system. Specifications for the data systems include the choice of sociometric, econometric, and psychometric measurements to be made on defined groups of observation units, and the instruments chosen or designed to make those measurements. This Operations Report specifies the design of the data entry system for the sample data systems.

Data entry design must take cognizance of the fact that various kinds and amounts of data will become available at different times for different groups of observation units, gradually building up fully longitudinal records in the MISOE sample data files for analysis, with all parts connectable for analysis and retrieval. The general purpose of the sample data entry system is to connect previously specified instrumentation (and groups of observation units) with finally usable data files. These final or master data system files are regarded herein as part of, as well as the goal of the data entry system, and therefore, must be specified in considerable detail with the file identification-descriptor system designed to ensure their interconnectability.

The definition of populations of interest is regarded as a part of the analysis specifications; the sampling of those populations as part of the specifications of the sampling data systems. The derivation of weights for data analysis, though regarded as part of the data entry system, is being specified externally to this document; the use of the weights is part of the analysis specifications. Thus, specifications for the data entry system include:

1. Design of answer sheets and control forms for optical scanning.
2. Logistic issues in data collection.
3. Optical scanning specifications leading to the emplacement of raw, basic data on magnetic tapes.
4. Tape editing and reformatting operations to convert information on the original tapes to tapes containing edited data in proper format for initiating, and then matching and merging with, the appropriate MISOE data files ready for analysis.
5. Special operations designed to ensure file confidentiality.
6. Special operations designed to ensure interim file connectability with previously developed files by match-merge operations in #4, above, and the interconnectability of the MISOE data files ready for analysis in the analysis system at any stage of their longitudinal development.

It will be convenient to proceed with the task of specifying the data entry system by making explicit some general assumptions (specific assumptions to be stated where required in the relevant context), to specify the MISOE sample data files and their identification-descriptor system, as detailed goals of the other operations in the data entry system, and then to specify how MISOE data entry operations are to achieve these goals.

Some of the general assumptions for specifying the SDS data entry system have been stated in the previous discussion of this system's place in total MISOE. An additional and important assumption is that initial processing of information will be done by optical scanning using an OPSCAN-100 (Digitek) and 24K Honeywell computer system readily available to MISOE staff.

Other data entry operations and analysis involve an IBM 360 computer facility in a different location. Specifications discussed in this document presume this hardware configuration, and might require adjustment at certain points if this situation changes. Given the anticipated volume of initial data processing, and the variations in data-gathering operations discussed above, the desirability of optical scanning rather than verified keypunching operations (which are generally less reliable) is enhanced. The choice of this particular scanning facility on the basis of ready availability requires special attention to error detection and control. In the event that this facility has such capabilities, they should be ascertained, reviewed, and used with such supplemental controls as may be necessary to ensure highly reliable document-to-tape conversion. This matter will be discussed in further detail in a later section on optical scanning. In this connection, it is noted that another facility using the same OPSCAN-Honeywell configuration is anticipated to be used for certain operations designed to maintain confidentiality, with separation of name and address from data files. It follows that similar error detection and control considerations will apply. These remarks and the more specific remarks in Chapter III should not be taken as criticism of the capabilities of these facilities, which are assumed to have been set up for generally simpler tasks with smaller volumes and less variety in specifications. The accuracy of these operations is extremely critical to the success of MISOE.

Other general assumptions are that:

1. All instrumentation has been completely and finally specified for first generation implementation of MISOE; any content, or format changes in such instrumentation, except for deletion of a whole instrument will require selective changes in the data entry specifications. Master Identification Form(s) cannot be deleted nor can the Input Battery Cover Sheet without drastic revisions not

only in the data entry system but to the connectability and confidentiality bases on which entire MISOE rests.

2. The operational MISOE data files as specified in Chapter II are final with respect to their number, plan for updating in longitudinal development, general layout, and exhaustive of the types of files for which sample data entry operations are required (as distinguished from any further files that may be derived in the analysis system or otherwise specially derived, and as distinguished from CDS files). Any expansion of MISOE involving violation of this assumption implies an expanded, or at least modified data entry system.
3. The SDS data sources and observation units have been completely and finally specified.
4. CDS as designed externally to this document has sufficient communality down to the program level in its file identification-descriptor system to permit connectability with the SDS files. This is required to enable movement of certain economic data from CDS to SDS and for certain anticipated weighting operations.

In Chapter II the general specifications for operational MISOE SDS files interfacing the analysis system will be presented, with the file identification-descriptor system designed to ensure connectability of MISOE components. The specifications for the individual file types as they are developed longitudinally will also be given. In Chapter III the general considerations for moving from available instrumentation to operational MISOE SDS files will be presented using flow charts and accompanying text to provide an overview of the SDS data entry system. More precise specifications of optical scanning and of the data entry operations for individual pieces of data entering the system will be presented in subsequent chapters.

II. SDS Files and Their Identification-Descriptor System

Introduction

The SDS files consist of the master files that constitute the goal of the data entry system and its interface with the analysis system, and the interim files that carry optically scanned information through editing, checking, and confidentiality-controlled identification to the initiation and development of the master files. The master data files are of five general types interconnected by a file and record identification system and a "link" system designed to protect file confidentiality. These five file types, to be described in considerable detail in later sections of this chapter, are:

(a) The "Program" file
(b) The name and address files
(c) The Student data file(s)
(d) The Teacher data file
(e) The Administrator file

In addition to these files, there is a special cross-sectional file which permits early analysis of student impact data, pending the development of the fully longitudinal student data file, and a "scramble" file; these will also be more fully described. All other operational MISOE data files will be regarded as "derived" files. Interim files will be described in later chapters.

Within a given full cycle or "generation" of MISOE, we shall consider each master file to exist once and only once (except for backup copies) in some stage of longitudinal development from initiation to completed file. Thus, we shall consider two kinds of "updating": (1) appending an additional section of the file record as longitudinal data become available, and (2) adding new groups of records as new cohorts come in at different time points depending on program length. This implies that records for a retired cohort are retained in the master files until a new MISOE generation cycle is initiated, thus permitting comparisons of aggregate information on adjacent cohorts in short programs, and preventing undue proliferation of master files. Records for observation units from different cohorts will be distinguished by a cohort number code that is part of the identification-descriptor system. The fact that a full master file contains records for various subgroups (different types and levels of students, teachers or administrators in different settings, e.g.) poses no problem so long as the identification-descriptor system codes are adequate. In the case of the high-volume, long-record student data file which in complete form is almost certainly a multi-reel file, only portions of which will be needed on disc storage at any particular time, it may be convenient to identify certain subgroups with certain reels and to regard these reels as subfiles of the master file. This approach provides a common basis for staff communication and interaction, even if it is decided to keep the master files on large disc-packs with the magnetic tape files kept as backups.

The Identification-Descriptor System

Unlike most simpler data systems, where random, serialized, or otherwise arbitrary file and record identification numbers suffice, MISOE requires an ID system in which at least some digits have substantive meaning for data processing controls and for analysis. We therefore refer to the system as an identification-descriptor system, because it describes certain types of files and records, and permits interfile connectability. This system, shown schematically as the leading portion of a MISOE tape record in Figure 1, consists of two major components: common and unique.

COMMON DESCRIPTORS						UNIQUE ID
MISOE Generation Number	Cohort Number	City-town code	School Code	Program code(s)	Grade/type code	IFID or PID

Figure 1. Schematic Layout of Identification-Descriptor Section of MISOE Tapes

One component is common in type and meaning across the types of data files but some of its subcomponents may carry different code values within a file. This "common component" consists of a 1-digit MISOE Generation number, a 1-digit cohort number, a 6-digit LEA number (3 digits each for city or town and for particular school), one or more program codes as described in Appendix A, and a 1-digit code indicating grade level and type of student involved. The first two digits (MISOE generati[illegible]ort numbers) will be serialized and start with "1". The LEA codes are [illegible]fied in documents of the Department of Education.

The grade level-type code is:

1-4 for secondary school grades 9-12, respectively

5-[illegible] for grades 13-14, for postsecondary programs at the community colleges, respectively

7-9 for adult programs (8 and 9 may not be needed).

The unique component consists of an arbitrary, but serialized number unique with respect to the observation units within a series code. For student and supervisor rating data, the anticipated volume indicates that this number have 5 digits (maximum allowance of 99999 is more than needed, except that total student volume may exceed 9999). For teacher and administrator data, three digits should be sufficient. However, it may prove to be more convenient in data processing for the complete identification-descriptor field to be of constant length throughout the system, in which case leading zeroes should be placed in right-adjusted unique numbers. Initially, these numbers, with the series code will be assigned at data collection time and called the initial file-identification number (IFID), which for confidentiality control will be replaced in certain files at specified processing points by a randomly-scrambled number, called the permanent identification number (PID). The confidentiality system will be described in the next section of this chapter.

It remains in this section to delineate the processes by which the identification-descriptor components get into the system. To do this, we need to anticipate certain features of data collection logistics to be more fully described in Chapter III. Each battery or instrumentation packet begins with an optically scannable "cover sheet" for the respondent's name, address, date of birth, and identification number. The latter includes the full identification-descriptor code with the common portion to be filled in by the respondent under the administrative directions, with the unique IFID precoded in the dark-mark field of the answer sheet. The purpose of the cover sheets is to initiate the name and address files and the associated confidentiality control system. Each packet will then be followed by a Master Identification Form (MIF) containing at Opscan time the full identification-descriptor code with sex and age codes to initiate and control interim data file processing for the entire battery. All instruments in the batteries including cover sheets and MIFs will be precoded with the IFID "dark-marked" on the optically scannable answer sheets. In the case of the MISOE generation number, it should be noted that when a second generation cycle is initiated, longer range followups of first generation students may still be going on, and may even overlap (in time) initial followups of second generation students in short programs. Without the cohort number, problems may be encountered with the IFID-PID system at cohort replacement time or during impact space operations with replacement cohorts, because the IFID-PID series would have to be continued, not restarted, across cohorts within a generation cycle. With the cohort number, IFID-PID numbering can restart with each cohort replacement, providing greater safety and flexibility.

The Confidentiality System

Confidentiality of information given by and about individual persons is a general MISOE requirement. This is met in the data files by numerical coding of responses and respondent identification. In longitudinal programs the need to contact some responding units at later points in time requires maintainance of name and address files. Confidentiality requirements can be met only by keeping such files physically separated from the data files with different identification numbers that can only be linked under very restricted conditions. The actual operations needed to set up and maintain confidentiality are part of the data entry system. A brief statement of the operating rules follows:

1. The name and address files will be prepared from the initial battery cover sheets by optical scanning by an external agency, called the link agency.
2. The link agency will have certain scanning, computer, and other data processing capability in addition to such professional qualification that permits its role in the confidentiality system under an agreement with MISOE.
3. The link agency will maintain the name and address files with back-up copies elsewhere, but not at MISOE.
4. The link agency will prepare a "scramble" file by assigning to each IFID from the cover sheets a unique random number of the same number of digits, called the PID. The scramble file will consist, except for a header record, of records containing the common descriptors, the IFID-PID pairs, and nothing else. The header record shall consist of the MISOE generation and cohort numbers and identification of the associated name and address file. Thus, there will be a scramble file for each name and address file (specified in the next section). Cover sheets will be required for confidentiality control and link agency processing for initial data collection batteries,

i.e., for student input, initial contact with teachers and administrators. For "over time" measurements on teachers and administrators, and process and product measurements on students, initial descriptors and IFID numbers are applicable (except for "gains" to be described below). Special handling of this problem for student impact measurement will be required. As a minimum, a cover sheet is needed for the supervisor data from the Massachusetts Job Evaluation Form.

5. The scramble file is the only file to pass between MISOE and the link agency.
6. MISOE will perform replacement operations (IFID to PID on incoming files; PID to IFID at followup time).
7. The link agency will optically scan the battery cover forms on which respondents will indicate name, address, and date of birth, the cover forms having been precoded (dark-marked) with the unique portion of the identification number (IFID).
8. At followup time, the student PID file will be derived and sent to the agency for IFID conversion and preparation of Impact Inventory OPSCAN sheets dark-marked with IFID, and preparation of mailing labels. Mailout will be performed by link agency. Returns will be received by MISOE.
9. MISOE input and interim files will contain IFID numbers as unique descriptors up to point of initiation of or merge with master files. Master files will contain PID as the unique descriptor.

The Name and Address Files

Name and address files will be maintained for students, teachers, administrators, and for supervisors named by former students to receive the Massachusetts Job Evaluation Form. In the case of the student name and address file, a new file will be created at cohort replacement time, starting with the new cohorts in the shortest programs, new cohorts from the longer programs being added to the "cohort file no. 2" as they come along. This implies doing likewise with the associated scramble file, both operations being done by the link agency. A whole new set of name and address files will be reinitiated with a new MISOE generation cycle.

The student name and address (and scramble) files will have one record for each person. Those files for teachers and administrators may have multiple records as described below. Beyond the header label, the name and address file records will have the schematic layout shown in Figure 2, and will consist of:

1. The full identification-descriptor code with IFID (five digits) in the unique portion
2. The name fields, separately defined for title, if any (6 positions), first name (10 positions), middle initial (1 position), and last name (12 positions)
3. The address fields, separately defined for apartment number or other qualifying designation (6 positions), street or rural route number and name (20 positions), city or town post office designation (12 positions), state (official 2-position alpha code), ZIP code (5 positions)
4. Date of birth fields (2-position fields each for coded month, day, and year)
5. A longitudinally developed section for coding mailouts, returns, and information such as refusals, deceased, etc., that may result either from followups of students in impact space or "over time" recontacts of teachers and administrators.

In the case of the supervisor rater files, the same person may conceivably be named by more than one former student. This will generate duplicate records which shall be distinguished by adding to the file record layout a field

Identification-Descriptor code	Title	First name	Middle initial	Last name	Apt. # etc.

Street address	Post Office	State code	ZIP code	Month of birth	Day of birth	Year of birth

Longitudinal record of mailouts, returns, deceased, etc. codes

Figure 2. Schematic Layout of the Name and Address Files

of five digits for the IFID of the student who named the rater. This has the further advantage of always having a recoverable record relating raters and ratees. The cohort number in these files should be that of the student ratee.

Respondents to any followup contact shall be asked in the followup instrumentation to correct and/or update their name and address labels. Some updating information is usually obtained about nonrespondents even without special efforts; where special effort is made to followup nonrespondents, additional information may be obtained whether the subject finally responds or not. For example, we may find that a subject is deceased. Therefore, provisions need to be made for updating the name and address files in the light of this information, including some that may become available at the school level about former students. In addition, at file update periods, additional fields should be added to the file records to code whether or not subject responded, was learned to be deceased, requested not to be contacted again, etc.

When preparing the name and address files, especially on students, printouts of the files should be inspected for obviously fake names and/or addresses and a delete code added to the file record without actually deleting the record, so that we can account for a lower mailout count than the initial counts. An example of a "goofy" record is: Elvis J. Presley, 99 Sunset Strip. Some are suspicious, but not necessarily false, e.g., Richard

M. Nixon (unless accompanied by 1600 Pennsylvania Ave.). The delete code may have a separate value so that an initial followup may be attempted with such suspicious names or addresses; this combined with the return codes will be useful in deciding the value of further followup attempts.

The "Program" File

The functions of the "program" file are to define and carry the common descriptors that interconnect the other files, and to carry the expenditure data collected from SDS-2 programs and capital expenditure data as allocated from CDS information. This connects the economic and noneconomic data in SDS. For SDS-1 programs, the special expenditure data fields will be legitimately blank. It is recommended that the capital expenditure from CDS be transferred for all SDS programs; if only for SDS-2, these fields will also be legitimately blank for SDS-1 programs.

In addition to the above information, the program file carries the stratification cell numbers and weights to be appended to other files selectively so that the analysis system can produce aggregate estimates of population parameters from sample data.

There will be a record on this file for each combination of common descriptors arranged from high to low in the following hierarchy:

MISOE generation number
Cohort number
City or town as coded in 3 digits
Particular school as coded in 3 digits
One or more program codes being developed by staff for CDS compatability and resolution of the multiple and cluster problems
Grade-level-type as coded in 1 digit.

A program given at more than one grade or level or to both adult and nonadult students will have multiple records on this file. A special and unique program code should be used in this and all other files to designate non-OE records on the files.

The program file will be initiated by the processing of the Student Master Identification Form (SMIF), administered to students at the beginning of their input battery. The interim file as edited and checked will be sorted on the hierarchy of common descriptors and a subfile prepared consisting of the first record and any subsequent record in which a single common descriptor varies from the previous record. To this subfile will be appended at process time the expenditure data in dollars by year and line item, allocated capital expenditures when available from econometric operations in CDS, and stratification cells and sampling weights as soon as available. The line items of expenditures should be placed together for the first years, followed by those for the second year, etc.

It is necessary to recognize that the initiating operations can only post program codes that are available at student input time as part of the descriptor system. It may therefore be necessary to add further information in process space (or post additional records) to permit handling of program shreds and clusters.

Although no plans or obvious need for IFID/PID information obtains for the "program" file, it may prove quite useful and convenient in subsequent data processing operations to post the ranges of these unique IDs on this file.

The Student Data File

The longest file in terms of number of records and the file containing the longest records per observation unit is the student data file. There will be one record in the master file for each student of any type in any SDS sample. The record is initiated by the processing of the student input battery. The first item in this battery is the "cover sheet" providing the basic information for initiating the name and address file and link system development. The second item in the battery, the Student Master Identification Form (SMIF) provides the common descriptor information to be posted in the same order as in the "program" file and at the beginning of the student file.

The schematic layout of the complete and final student data master file is shown in Figure 3. When editing the SMIF interim file, the MISOE generation and cohort numbers will be program added to the descriptors field preceding the city-town code. The SMIF also provides the IFID, which will be posted immediately following the common descriptor section at the beginning of the file. Data on the sex and age of the student, picked up from the MIF will start the input data portion of this file.

The complete identification section of the file will be immediately followed by fields for the stratification cells and weights to be posted as soon as available but before initiating the Master file. The file will next contain the scores from the input battery. The order of these scores is not critical, except that the sex and age information from SMIF should come first and the MPI data kept together in item order last. The standard instrument scores may be conveniently placed in order of administration with all scores from a given instrument kept together.

ID and Weights Section:	Common Descriptors from SMIF	Unique PID	Stratification cell and weights	Sex, age, and race from SMIF			
Input Battery Section:	ITED Scores	DAT Scores	Study habits and attitudes	SPV SIPV	HSPQ Scores	Culture-fair scores	MPI section
Process Battery Section:	School Sentiment Index Scores	Attitude Toward Program	Enrollment & Attendance data from schools				
Product Battery Section:	Completor-noncompletor	ITED retest	Ratings on Terminal Objectives	Information on noncompletors			
Impact Battery Section:	MEII Data 1-yr. later	MJEF aggregated data, 1-yr. later	MEII and MJEF macrofields in same arrangements for later followups				

Figure 3. Schematic Layout of Student Data Master Files

The master student data file will become longitudinally developed as process, product and impact data become available and are processed, from administration through scanning and editing, to merge point. Further details of these processes for the various elements are delineated in later chapters of this document. What is crucial is that the unique portion of the student identification be consistently dark-marked and identified with the same student, uniquely, in all subsequent data gathering processes, and carried through the optical scanning and interim file operations to point of merge with the previously existing master file. Similar considerations apply to the teacher and administrator data files described below.

The complete tape specifications including parity, track size, density, blocking factor, and tape position numbers with record volume and external label title must be developed from these general specifications and documented. This is a general rule throughout the data entry system for all interim and master tapes, data, name and address, or scramble tapes regardless of stage of longitudinal development. These should be prepared by data processing personnel as the tapes are developed and a library record tape kept, which contains this information, except for position layouts, and whether tape has been scratched, number and location of backup copies. If this is not initiated and faithfully maintained during operational MISOE, the need for it will become increasingly apparent and catchup will be very difficult.

In the case of the student data master file, no process data and limited product data will be obtained for SDS-1 programs, and the file can be split at any time between completion of input and merge of initial process data with the SDS-1 file having a shorter record. This has the advantage of faster and less costly loading and other operations at the interface with the analysis system when dealing with only SDS-1 or SDS-2, a gain that will not be apparent until nearly complete longitudinal development in SDS-2. If both master files are to be kept in internal disc storage, there is no obvious advantage to the separation. The alternative, then, is to keep the single master file with many legitimately blank fields in the SDS-[illegible] records of constant length. If variable length files are convenient to work with, they may be considered as a third alternative, especially after disc loading.

The Teacher Data File

The teacher battery may be entirely regarded as in process space and is entirely in SDS-2. The general file layout consists of the identification-descriptors, stratification cell of the sample school in which the teacher is employed, relevant weights, and the teacher battery data. Some measures, are replicated over time during the life of a student cohort.

This file consists of relatively short records, one for each teacher of a program cohort. Thus, the same teacher may have multiple records, a new record being generated at cohort replacement time for any program in which the person teaches, or separate records being generated for each program taught. Those measurements which are not replicated over time do not have to be readministered at cohort replacement time, so that the original scores can be moved to the added record at merge-to-master-file time. A similar principle applies at file initiation time where it is unnecessary for the same teacher to take any instrument more than once, except for the Attitude Toward Program, if that person teaches more than one program. Even with the relatively small battery and numbers of teachers involved, this should represent some savings in cost and considerable savings in teacher morale as a function of their participation in MISOE.

It will be convenient to order the measurements on this file by first placing the nonreplicated measurements such as IQ, followed by those replicated measurements given together at $time_1$, $time_2$, etc., respectively.

Since header labels will distinguish the data records on interim teacher files from those on the other files, the IFID-PID numbering system may be restarted from 00001. Note that the initially administered teacher battery will include a Teacher Master Identification Form (TMIF). Sex will be moved to the nonreplicated measurement field, age to the replicated.

It remains to deal with the gain, loss, or transfer of teachers during the life of a student cohort. If a new teacher enters a school and teaches any relevant SDS-2 program or non-OE control group, or is shifted into this situation either from SDS-1 programs or from other schools (sampled or nonsampled), that teacher is to be considered as a "gain". At the next "over time" contact with that school for replicated measurements in the teacher battery, "gains" will receive the cover sheet, the TMIF, the IQ test, and the full battery. The cover sheet will generate an addition to the name and address and scramble files for teachers at the link agency. A new record will be added to the teacher master file with TMIF and IQ information in the usual positions, but with the replicated measurements in the proper time position in the "overtime" replication fields. The replicated measurement fields for initial and any intervening contact points will be left legitimately blank. The remaining record generated by later replication times will be generated in the usual manner unless the person becomes a loss. It is recognized that the additional record from a transfer may be a duplicate (with respect to the individual) on the data files, in which case the original record should from this point have a loss pattern, but be retained, because all other information, except certain descriptor codes will be different. It is assumed that no administrator, unless definitely demoted will become a teacher in which case his addition to the teacher files will be accompanied by a loss on the administrator files. In the case that an administrator is temporarily teaching to fill an emergency situation, no such gain and loss operations will be carried out. It is also assumed that, despite the fact that teachers sometimes have administrative duties, they will never be regarded as administrators unless definitely promoted to same, in which case gains and losses will be treated across the teacher and administrator files accordingly.

A "loss" is defined and treated in exactly the reverse manner. The original record becomes legitimately blank from loss point on unless regained, but it is retained rather than deleted. A new record on the teacher files is generated, if the loss is really a transfer from one SDS-1 program to another or from one school to another with an SDS-2 teaching load (or non-OE control); a new record is generated on the administrator file if the "loss" of a teacher is really a promotion.

The Administrator Data File

Although fewer measures on fewer subjects are involved, similarity to the teacher file is striking. The similarities in generating this file and the name and address, scramble, and data files for administrators are as follows:

1. It is possible for the same administrator to have more than one record on file, with some different common descriptors.
2. All measurements are process measurements in SDS-2.
3. Some measurements are replicated over time; some not.
4. General ordering of file components is exactly analogous.
5. The header labels on interim files will be distinctive and the IFID-PID numbering may be started at 00001.
6. Initial administration of the Administrator battery will include the cover sheet with subsequent operations at the link agency generating the name and address and scramble files, and will include the Administrator Master Idnetification Form (AMIF). Sex and age will be moved to their appropriate measurement fields.

7. Gains, losses, and transfer rules are the same in general with some detail specified above under that for teachers.

The Special Cross-Sectional Impact Files

Rather than wait for maturation of student cohorts through the longitudinal process from program entry to time points beyond graduation, where they are in impact space, to be able to do any impact analysis, it has been decided to obtain some impact data on former cohorts on a cross-sectional basis at initial implementation of MISOE. Because little or no input, process, or product, or economic information related to the process will be available, there will be no connectability and analysis will be confined within the impact space. This implies certain ad hoc data entry operations with very limited application to those operations for the longitudinal impact space. Moreover, the possibility of one-year, three-year, five-year, and/or 10-year cross-sectional impact samples, which are not connectable by individual implies separate name/address, scramble, and data files for each followup lag. Moreover, name and address files for the students will not be used again for further followup, so they need not be maintained through the link system, but must be processed in such a way as to maintain other features of the confidentiality system. The supervisor names and addresses supplied by student respondents must also receive confidential handling. Because the special aspects of this whole operation have not been completely specified, certain explicit assumptions and suggestions will be presented here as a basis for discussing the development of the master files for the cross-sectional samples.

1. We assume that local schools can and will provide as a minimum information about former students identified by program, date of entry, date of exit, completion status, name and last known address, and possibly sex and date of birth, and nothing else. It is quite possible that different schools

will have different capabilities with respect to the form in which they can supply this information. MISOE staff must therefore be prepared to convert the multiform supply of this information to some common base to be placed on tapes, possibly even supplying the clerical personnel to place it on scannable forms or keypunch and verify it. Not all of the common descriptors will be available. The MISOE generation number should be assigned as "0" and the cohort number reflect the followup lage, "1", "3", "5", and "0" for the 10-year if included. The LEA and program descriptors should be available, but grade level of entry possibly not. IFID's can be assigned 00001 upward "pseudo-cohort" number.

2. Given that lag cohort files with this information and descriptor system have been prepared by MISOE, mailing labels can be prepared and the appropriate impact instrument mailed out. The file can be sent to the link agency for preparation of a scramble file, and a name and address file for more normal operations in the confidentiality system (for following up nonrespondents and incidentally to get the name and address file with IFID numbers out of MISOE before mailed questionnaires are returned with IFID numbers on them).

3. Respondent data could be treated in the normal manner for impact space and the supervisor name and address file prepared. The specifications for normal impact space operations will be delineated in a later chapter.

4. Data files for respondents and supervisors could be processed in the normal way except that they initiate corresponding cross-sectional master files rather than develop longitudinally already existing master files. Master file structure would presumably consist of the descriptor section with PID and normal impact sector structure.

III. From Instruments to MISOE Master Files

Introduction

With the previous chapter describing the Master files and other permanent files as the goal of the data entry system and as interface with the analysis system, some glimpses of the vast territory between instrumentation and final files were presented. We must chart that intervening territory in much greater depth and detail. Given instrumentation, we must discuss the intervening processes in four phases: (1) data collection, (2) optical scanning to initial tape production, (3) operations on input tapes leading to final stage interim files, and (4) operations on the latter that initiate or develop the Master Data files. These four phases will be described in terms of the student instrumentation to student master files, and down to a level excluding the detailed scoring and editing specifications, which will be taken up in subsequent chapters along with other data entry details, such as special considerations of the teacher and administrator data operations. The reader is encouraged to consult the table of contents as a guiding outline of the data entry system as detailed in this and subsequent chapters.

The Data Collection Phase

One or more instruments will be administered together, as a "battery", to diversely located responding groups over a time period. Response to each instrument will be made on one or more answer sheets optically scannable by the OPSCAN-100 system. The entire set of scannable answer sheets will be placed in order of administration of the instruments in a "packet", with each answer sheet precoded in the dark-marked area with the same IFID number. At the head of the packet for initial administration of a battery to students, teachers, administrators, and "impact" supervisors, there must be a "cover sheet" with the common IFID for that packet and alphanumeric grids permitting respondent coding of name, address, date of birth, MISOE generation and cohort numbers. Also on initial administration there must be a Master Identification Form, as the first instrument, with its associated answer sheet. With both the cover sheet and the MIF, the answer sheet and the instrument may be physically identical, although logically distinct. The MIFs will also contain the MISOE generation and cohort numbers. In both cases these may consist of a single circle that the respondent must fill in, and in future operations, where these will not automatically be read as "1", editing specifications for the input tapes will convert them to the proper values. This will be of no concern in the first MISOE generation cycle for the MISOE generation number; as each new program cohort comes along, the editing specifications will be a function of the header label on the input tape.

Instruments consist of those commercially available with scannable answer sheets already available, those commercially available but without system-compatable answer sheets, and those built by MISOE to meet its own requirements and for which answer sheets must be designed. We need to look at these three situations somewhat closely.

1. Commercially available with scannable answer sheets: Arrangements must be made with the test publisher to have the IFID precoding in the dark-mark area (5-digits in a six-digit field, right adjusted preferably, but if not, the 5-within-6 positions must be consistently applied and documented to MISOE). A general specification of MISOE's data entry at this point is that scanning fields be defined by individual item response-alternative positions leading to dichotomous reading as "1" or "blank". This decision was made to provide maximum flexibility for editing specifications and generated variables at later stages of the data entry system, and is most applicable to the MISOE-developed instruments. However, the predesigned answer sheets which accompany some of the commercial tests may have fields defined at the item level rather

than at the response-alternative level. This matter should be determined immediately so that necessary adjustments in scoring and editing specifications can be made. MISOE should go with this rather than redesign special answer sheets.

2. Commercially available without scannable answer sheet: This would seem to apply primarily to the Leonard Gordon "values" instruments by SRA. Here, MISOE must design the answer sheet, obtaining publisher permission to place the instrument on, or attached to, the answer sheet to ensure item to response field coordination.

3. MISOE-designed instruments: Answer sheets must be designed ad hoc with the response-alternative level of "field" definition and close coordination of instrument and answer sheet design ensured. For this, certain flexibilities such as physical attachment of booklet page and answer sheet by perforation, or in the case of very short instruments, printing items on the answer sheet should be used.

Since the optical scanner requires a "control" form that defines the location of response fields to be read, these controls must be perfectly coordinated with the answer sheets. In the case of the commercially available answer sheets, controls may be presumed to exist or be readily created from detailed examination of the response locations on the answer sheet. Where answer sheets must be designed by MISOE, control forms should be created in conjunction with the answer sheet design to ensure intercompatability of instruments, answer sheets, and scanning controls.

MISOE-created answer sheets must be printed in quantity and it is very important that printing contracts specify absolute adherence to space tolerances to ensure that completed answer sheets are scannable (response positions must line up with electric eyes detecting reduced reflection of light when positions are marked). Otherwise, massive and systematic errors will occur in the data.

The logistics of collating data collection materials, disseminating and collecting them, and the instructions to battery administrators and respondents are all critical matters bearing on the reliability of MISOE data. First-contact batteries such as student input, first-time administration of teacher and administrators batteries, should be closely monitored at the local testing sites by MISOE personnel. In the case of teacher and administrator batteries, consideration should be given to the idea of MISOE personnel being the test administrators. In the case of the student input battery, preliminary seminars for teachers who will act as test administrators should be conducted by MISOE staff to ensure as much uniformity of data collection procedures as possible. In some parts of MISOE, e.g., in the followup instruments in impact space, complete self-administration must be tolerated. In all cases of multi-instrument batteries, instruments and answer sheets must be coordinated and answer sheets collated by IFID. In the student input battery which is sectioned over several testing sessions, the whole packet of answer sheets must be passed out and back with the cover sheet intact so that the same student receives the same answer sheet packet back with the same IFID at the next testing session.

At the beginning of the last testing session, the cover sheets should be removed by the respondents as soon as packets have been distributed, and collected by the administrator, who will put them in a box or manila envelope in the presence of the students. This package will be addressed to the link agency with the return address of the school. MISOE will know the presumed date of this by school and the link agency should log in receipt of these materials for checking the transmittal. The answer sheet packets and any reusable test materials should be collected at the end of the last testing session for return to MISOE. (The answer sheets may go directly to the optical scanning facility with proper logging controls.) It is recommended that #2 pencils be supplied by MISOE to encourage uniform usage of the proper marking consistent with optimal scannability.

Since the MPI is the most intrusive of the instruments in the student input battery, its administration on the last day after cover sheet removal is recommended. While this may be contrary to previous scheduling specifications, serious considerations should be given to this issue, before final scheduling and other operations dependent on the testing logistics, like preparing administrative instructions.

Instructions for administration to the test administrators and to examinees must be prepared and they are critical reliability controls. Hence, the care and thoroughness of their preparation cannot be overemphasized. Some guidelines for preparing these instructions follow:

1. For the commercial tests, the administrative manuals and any test instructions to examinees should be carefully followed. In fact, they may also be used as form guidelines for preparing instructions for use with MISOE-constructed tests. All testing times and logistics specified should be adhered to as closely as possible (but see #3 below). It is recognized that those tests with modified answer sheets or connections between answer sheets may have to be partially revised. All commercial manuals for administration and instruction to examinees on testing materials should be carefully reviewed for consistency with the MISOE logistic requirements.
2. Instructions for completing the cover sheets and MIF's must ensure that they are filled out completely without exception. It is at this point on the very first day (to be reinforced at each testing session) that general instructions be given to examinees for handling the optically scannable answer sheets. Most students have presumably used them before, at least in the larger school systems; they may be a new experience for some students in smaller or more isolated schools.

3. Most instruments in the battery are "power tests", timed to ensure completion by about 95% of the students. Unless field testing has indicated higher rates of noncompletion than 5% within recommended time limits, these recommendations should be followed in the case of the commercial tests. Clerical Speed and Accuracy tests are deliberately timed to make completion unlikely. Close adherence to the recommended time is very important in this kind of "speeded" test. Some space tests are also speeded to make responses less dependent on reasoning. MISOE should consider obtaining and supplying to examination sites a set of prechecked stopwatches; if this is not feasible, those available at local sites should be carefully checked by test examiners.

In the case of MISOE-developed instruments, especially inventories, the timing should be based on whatever information field testing gave about time to complete, and some judgement made about the limits to be specified such that 95% complete. Where this information is uncertain or based on superior subjects, it is better to err on the side of allowing more time. Valuable information may be systematically lost from the end of the inventories if too many fail to complete them. Missing data imputation may then be based on a somewhat more biased group of respondees to these items than to others generally, using the modal response method, and greater risk is involved in a priori imputation if it has to be applied to too many subjects.

Attention must now be given to the situation in the student input battery where a student does not report to all testing sessions. Assuming that this will be relatively infrequent and sufficiently sporadic to be regarded as random loss, no attempt should be made to recover the first day or to administer subsequent days' testing for those who miss the first day. Except for all of this first day being missed, the records for those who miss parts of the battery (late, taken ill, etc.) should be included in the MISOE data system, unless more than a whole testing session is lost. Even those who miss some of the first day, but get through the cover sheet and SMIF should be retained if they complete the rest of the battery. Persons missing the last session of testing but completing prior sessions will not be able to remove the cover sheet; this could be done by the test examiner. Only in the case of catastrophic or epidemic losses should makeup sessions be scheduled. A major catastrophe resulting in loss of a whole school or program is likely to be unrecoverable and first-stage weights will have to reflect this. Processing consequences of students missing whole instruments are discussed in Chapter IV.

The Optical Scanning Phase

Given that the batteries have been successfully administered and that cover sheets have gone to the link agency while all others have arrived at the optical scanning facility, we consider now the operations for quality control scanning of the answer sheets to the production of the initial data input tapes. Because item response positions are generally located on the answer sheet in such a manner as to coordinate with convenient response to the test instrument, and hence may be arranged both horizontally and vertically in the answer sheet matrix, while the answer sheets are scanned only horizontally, the item response data will be scrambled across items on the initial input tape. It is this scrambled tape, produced by the OPSCAN-100 with a Honeywell computer system in even parity that will, on quality control approval, leave the scanning facility and enter the MISOE IBM-360 data processing facility for parity change, unscrambling, and data editing operations.

A batch of answer sheets will come in for a particular battery and should be processed by instrument, even though some instruments may have more than one answer sheet. That is a more or less idealized presumption and we need to consider likely exceptions. Answer sheets for a battery will not necessarily arrive at scanning facilities at the same time, almost certainly not from various sites, possibly not by programs or testing groups within site, and even possibly not for all instruments within the battery (although some of these potential hazards may be minimized by careful transmittal instructions to the testing sites). In any case, it is absolutely essential that the scanning facility keep logs of input and processing of information.

The general plan for producing a single data input tape is to scan the answer sheets for a particular instrument administered to a particular group (e.g., secondary students) from all sites and across programs and grades. The final input data tape will consist first of the header label record specifying the instrument and group with MISOE generation and cohort numbers. This will be followed by the scrambled order of item responses to the first side of the first answer sheet, those to the second side of the first answer sheet, those to the first side of the second answer sheet, etc., in that order, and as relevant to the particular instrument. A temporary input tape should be initially produced for a single answer sheet side, and the multiple tapes sorted and merged to produce the input tape for the MISOE facility. This will ensure that sides and answer sheets will be collated on IFID number on the final tape. A similar principle of temporary multiple input tapes merged to a single instrument-group-oriented tape can be used to proceed with the processing of large batches without waiting for all testing sites to get all of their answer sheets to the scanning facility.

We now consider the quality control operations at the scanning facility. Despite the experience in these matters of the scanning facility personnel, who could no doubt improve on any suggestions here, a few points appear to be critical in the anticipated large volume, multi-source input, especially in the case of the student input battery.

1. At the log-in stage, quick visual inspection should be made of the order in which answer sheets have been submitted and careful separation by instruments made to define process batches. At this stage, a batch should be inspected for any gross or obvious nonscannability such as systematic use of a pen, or systematic light marking. These mini-batches may be combined across testing sites with those for the same instrument to form a processing batch. Subsequent control points refer to such a batch at the single-side-of-an-answer-sheet level.

2. It is of course critical that the proper scanning control form be matched with the answer sheet batch. It is assumed that an error here yields a false record, rather than a hangup. It is also assumed that normal scanning operating procedure has some control of this possible error source. If such an error occurs, it should be detectable by the quality control point #5 below.

3. An initial pass of the batch through the scanner should be made at normal operating sensitivity levels. Those falling in the reject pocket should be examined visually and carefully for the reason. Here we should rely on the experience of scanning facility personnel to define more precisely the error detection and correction procedures. Assuming these are successful, and the sheets are then scannable in one or more waves of this step, the batch is now on a tape.

4. A "missing mark" selection option available on the system should be used, if feasible, to cause the rejection of any answer sheet with missing marks for more than 5% of the control form response fields in commercial tests and 15% for the MISOE-generated inventories. Falling in the reject pocket, inspection should center on whether missing marks are sporadic or systematic, cluster at the end of an instrument, or cluster in the item-related fields where missing data at the item level are legitimate.

5. Answer sheets which were presumably read onto the tape should be kept in order of scanning within the batch and a tape printout produced. The first and last few records plus a 1% sample of all records in the batch should be visually checked on the printout against the answer sheets to be sure of successful tape production. In the case of the tape produced from the SMIF, this should be done within sub-batches coming from the different schools. Although answer sheets from different schools were presumably combined to define the batch, the answer sheets are likely to be together from a given school in the answer sheet stacks (within accept-reject waves) and on the tape. It is unlikely that sufficient program grouping within schools will permit this checking arrangement to be feasible at the program level. However, the 1% sample must be examined for successful identification-descriptor generation and transmission.

Because it is difficult to anticipate the nature and frequency of the various errors that might be encountered, it is not feasible to specify at this point what corrective actions must be taken. Experienced scanner facility personnel can be helpful here, the general guideline being to maximize reliability and validity of the operations leading to trustworthy input tapes. All problems encountered and their resolution should be documented for MISOE records. This documentation will be useful in improving operating procedures at cohort and MISOE generation replacement times.

The lower volume and variability in respondent behavior anticipated for the teacher and administrator batteries indicates that absolute error detection and control will be less critical than with the student input and impact batteries (both rater and ratee response). Nevertheless, quality control should be used throughout the data entry system.

Because the control forms are set up for field definition on the main rows of the scan-matrix with auxiliary rows automatically controlled by the field definitions in the prior main row, the use of the auxiliary rows should be minimized, and confined to certain response field patterns with constant field specifications. This needs to be taken into account in the design of the answer sheets.

With the exception of the header label and the contents of the name and address files, all input files will be produced by scanning numerical fields. Moreover, the presumption is that these fields will be defined by single item-alternative response possibilities leading to all "ones" and blanks on the input tapes. A previously noted exception is the case of any commercial instrument for which answer sheet design and control is inconsistent with this rule.

Sub-batch tapes based on individual sides of individual answer sheets may, when all have passed the quality controls, be merged at the instrument level for transmittal to the MISOE IBM-360 facility and editing operations. The detailed layout of the transmitted input tape _must_ be transmitted therewith.

Processed answer sheets should be retained either in the scanning facility or by MISOE until the information has passed through the rest of data entry and is on the appropriate master file. If archival storage facilities can be found at reasonable cost, all answer sheets except the cover sheets should be retained through a MISOE generation cycle. Ultimately the answer sheets should be shredded and burned. Those from the cover sheets should be destroyed in this manner as soon as the name/address file and its backup have been prepared and verified.

Operations on Input Data Tapes

The operations on input _data_ tapes will be carried out at the IBM 360 facility available to MISOE. These operations consist of a series of scoring and editing steps and involve several sets of ad hoc computer programs for their accomplishment. The information passes from the input data tape to master file through interim file tapes, which can be scratched as soon as the next file in the series is checked and documented. The header label of the input tape or of any interim file tape should be carried through to the next tape with minor change to indicate each interim file uniquely. All tapes in the system should also carry external labels and their nature and status should be carried in the tape library record file.

The first interim data file should be prepared by a program (type I) which reads the even-parity, Honeywell-produced, input tape, modifies the header label, changes parity, unscrambles the data macrofields associated with each answer sheet side, and outputs an odd-parity, unscrambled tape. It should also make any channel number and/or density changes that will make subsequent processing uniform and efficient. Nine-track, 800 bpi files may be optimal; data-processing personnel can decide this in light of their experience in working with a particular facility and in terms of the ease of using external facilities in the event of major breakdown or loss of the present facility.

Unfortunately, separate programs to accomplish this step, and those to be described for future steps, must be written ad hoc for each input file because of the unique relationship of the unscrambling and editing steps to each source instrument. The output file will have the same general format as the input file except for the unscrambling effect.

The second interim file should be prepared by a program which reads the unscrambled input file (interim file 1), performs certain scoring and preliminary editing operations, and outputs a file for final editing. The scoring and preliminary editing operations may be few and/or simple, or may be extensive and/or complex, depending on the instrument which generated the file. Detailed specifications for the program will be presented in subsequent chapters,for each major group of input files, instrument by instrument, and for the MISOE-generated inventories, item by item. In this section, we present the outline of operations that this program (type II) must perform with certain suggestions for programming efficiency. For example, use of variable formatting may make it possible for the interim file I/O operations to be generalized across programs of type II. Moreover, certain features of the data management package in coordination with the general flexibility of IBM 360 equipment should facilitate such programming, along with the use of ad hoc subroutines that may be called for the special requirements in processing a particular file.

Such subroutines must be designed to perform the scoring and preliminary editing operations for the specific instrument-oriented file. We will now specify these in a general way with comments regarding the files to which they are applicable, and certain general ways of finding the necessary input information in addition to the unscrambled file being treated. The detailed specifications in subsequent chapters must be consulted to code the type II programs in detail.

Scoring operations for the commercial tests consist of programming explicitly the commercial scoring keys for the instrument, possibly in tabular form, relating test item number and/or alternative printed on each key to their position on the unscrambled input file (interim file 1) tape. For each of these relations, the presence of a "one" or blank on the tape position enters the formula for a test score. Usually a single key will produce a single score, with multiple keys generating multiple scores. Some tests may be scored by a correction for guessing formula and have a "rights" key and a "wrongs" key so that the scoring formula, instead of being the cumulation of the number of "1" (or other item score on the tape) indications across the key, is such a number from the rights key minus such a number from the wrongs key with the difference divided by a constant. In some cases, as in a Cattell factor instrument, the factor weights are built into the key rather than applied after scoring to generate a factor score. Such weighted keys require the program to sum the weights of keyed items rather than summing unit counts. In some cases, the nonresponse or blank positions may be weighted, other positions on the key and tape being ignored as irrelevant to that score. In one of the keys, both of a pair of responses must be present to add to the score.

Scoring operations for the MISOE-generated inventories will generally consist of moving individual item "ones" and changing "blanks" to "zeroes" where this is legitimate, or reading a pattern of ones and blanks over several positions to determine item metric score. In some cases, it will be required to post an a priori imputed value where the pattern consists solely of blanks.

When these operations have been completed for a given record on the file, the program should then take these obtained scores to generate any additional variables that are functions of one or more of these scores, and as specified in detail in the subsequent chapters. Such additional variables from a given instrument are of two kinds. One kind is a metric or recoding change such as conversion of a raw intelligence test score to an IQ score, or any raw score to a prespecified standard score. To accomplish this, the program must read into storage or explicitly code the specified conversion table provided in the scoring manual for the test. The rest of the operation is a "table lookup". The other kind of additional variable is an algebraic sum, possibly weighted, of one or more of the scores obtained earlier in the program.

The last section of the program before reading output records to interim file 2 is to range-check each variable, whether originally scored or added as metric-changed or generated variables, keeping a count and printout of the IFID numbers and total count of such out-of-range scores detected. This printout should be retained as part of interim file 2 documentation. Out-of-range scores will be replaced in storage by the a priori imputed values where these apply as above, and by blanks otherwise. If more than 5% of any score is found to be out-of-range, the programming of the scoring keys or specifications should be carefully rechecked, and if not found to be faulty, the unscrambling operation should be reviewed. If necessary, the fault should be traced back to the scanning level. The completed records can be read out serially behind an appropriate header label.

With the successful production of an interim file 2, its layout, the range-check printout, and other pertinent information should be documented and the information posted to the library record tape file. The interim file 2 now becomes input to a program (type III) which imputes values for missing data in all score positions for which the a priori method used in program type II was inappropriate. Before executing this program, frequency distributions for all such scores should be run with the statistical package. These distributions will be inspected to locate modal values to be keypunched and verified for read-in to program type III. The specific items and frequency distributions specifications will be given in the subsequent chapters.

Program III can be written to read in interim file 2, and the cards (or tape produced from the cards), containing the imputed values and interim file tape positions for the variables involved. The body of the program detects blanks in these tape positions and replaces them with the imputed values, then outputs interim file 3 ready for merging and replacement of IFID numbers with PID numbers. The imputation deck should be printed out by the program as part of the documentation of file 3. File 3 format is the same as file 2 except that there is a minor change in the header label. All scores not affected by this missing data imputation step are simply moved from interim file 2 to readout area in proper loci, bypassing the imputation step.

Battery Merge and ID Change: Getting to the Master File

For each instrument in a battery we have built an interim file 3. For a battery there are several such files which should now be sorted on IFID number and merged. The merge program (type IV) should be written to read in the individual interim files (3) as sorted and output a battery file with a new header label, and with the records reformatted to the arrangement specified in Chapter 2 for the appropriate portion of the appropriate master file. When the battery file has been completed, checked, and documented, it is ready for conversion of IFID number to PID numbers using a copy of the scramble file from the link agency.

The battery file with PID numbers is now the master file in the case of an initial contact battery (e.g., the student input battery), and the header label should so indicate. In the case of a battery developed from some part of the longitudinal design of MISOE (over time replications and followups), the battery file with PID number must now be matched on PID number with the previously initiated master file. In this situation, the battery file header label will not indicate that it is a master file; the master file header label will show, after merge, the longitudinal status of its development.

It is recommended that battery files with PID numbers not be scratched, but retained as backup files permitting ready regeneration of master files if necessary. Battery files with IFID numbers can and should be scratched.

Debugging Operations

All programs written to carry input files to the master file stage must, of course, be thoroughly debugged. Moreover the chain of operations must also be debugged to ensure that the data on a set of battery answer sheets gets to the master file in proper form and content. It is recommended that about a dozen sets of answer sheets be completed by MISOE staff using IFID numbers in the 99900-99999 range, to be processed through the scanning and interim file set of operations and computer programs and the results carefully handchecked against the original answer sheets. In completing the answer sheets, a variety of likely patterns of missing data should be induced. This should be done for each battery with its groups of ad hoc programs written at the source instrument level. This extra effort to ensure that not only individual programs are valid but that the whole set of operations is validly connected will undoubtedly save MISOE some expensive grief, of which reprocessing a large amount of information in large volume may be only the most obvious and relatively minor example.

IV. Data Entry Operations With The Student Input Battery

In this and in the next several chapters, we will detail those scoring and editing specifications and the unique aspects of the interim file programs, postponed from Chapter III. Because the student data file development involves such a large set of batteries and instruments, the IPPI elements will be treated separately, starting with the input battery in this chapter. The more detailed specifications for these data entry operations for the process, product, and impact batteries will be presented in Chapters V, VI, and VII, respectively; those for the teacher and administrator batteries will be presented in Chapter VIII.

The special nature and reduced requirements for processing the cover sheet and Student Master Idnetification Form require the editing operations be done on the scrambled output file at the scanning facility. Thus, for these two exceptions, and in addition to the quality control operations previously specified at the optical scanning facility, there are no scoring, metric change, or generated variables, or typical missing data imputation operations requiring program types II and III in the IBM 360 facility. There will be subsequent operations at the link agency for the cover sheet file and at the IBM facility for the SMIF file. The scrambled scanner output file for both should be listed completely on the Honeywell printer; the records should contain no blanks but be entirely numeric and contain the full identification-descriptor data. For the latter, values other than "1" are expected and valid, except initially for the MISOE generation and cohort numbers. (In case the answer sheets have been designed in strict adherence to the unit-position field definition rule, "ones" will be expected and non-ones will be regarded as out-of-range. These will have to be converted in a type II program at the link agency or IBM 360 facility as relevant.) The printout at the scanning facility must be carefully examined for the completeness and range. Missing or out-of-range positions must be replaced by the imputation of the appropriate MISOE generation and cohort numbers, and in the case of the other fields, resolved by reference to the answer sheet stacks. It is for this reason that these operations must be done at the scanning facility while answer sheet groups are being processed.

The cover sheet heading the student input battery has been specified in the previous chapters in regard to functions and content. It was assumed that it consisted of a specially designed, scannable answer sheet (one side) on which the requested information was identified to the respondent, and that administrative directions would ensure completion of the form, which would be sent to the link agency for alphanumeric processing. We now move to the instrument-by-instrument specifications, starting with the Student Master Identification Form.

The SMIF

Like the cover sheet, this form consists of a specially designed, scannable answer sheet (one side) with requested information identified to the respondent. Again, the administrative directions must ensure completion of this form. The information from this form initiates the longitudinal development of the student master data file. It is reasonable to edit the sex (1 for male, 0 for female), age (9 dichotomies), and race (5 dichotomies) codes while this scrambled file is at the scanning facility and problems can be resolved by reference to the SMIF forms.

When the scrambled SMIF file has been transferred to the IBM 360 facility, it must be unscrambled by a type I program. The type II program can be skipped, unless data are all "ones", in which case patterns must be examined to write out proper identification-descriptor codes. The file, except for PID replacement of IFID numbers, is now the initiating student data master file. It should be held pending completion of the other interim files of type III from the student input battery for battery merge and PID replacement of IFID numbers.

There is another special reason for holding this file available in the IBM 360 facility (actually a copy which might be spoiled in the special usage about to be described). Some students may miss one or more specific instruments of the test batteries (came in late or were taken ill, etc.), in which case no incoming answer sheet or a blank answer sheet in their packet will be received. If blank answer sheets are received with their IFID numbers and are processed, the specific treatments for missing data in instruments, mostly commercial, where raw scores are generated by programming scoring keys will result in zeroes instead of blanks for the missing raw scores. If, however, no answer sheet is received (error in transmittal logistics, came loose or got lost) no record, not even the IFID, will be generated for the scores for that instrument, and at battery merge time, there will be various kinds and degrees of nonmatches. To head off, or control, these problems, the following actions must be taken in the type I programs for unscrambling, or as a step to be taken on the unscrambled files before entering the type II programs, in processing each instrument file:

1. Match the interim file 1 with the SMIF file on IFID number (both files presorted on IFID number). Printout nonmatching IFIDs for MISOE documentation.

2. For IFID's with records on the SMIF file but not on the interim file 1, generate a null record for interim file 1 with the missing IFID and add it to interim file 1. The null record consists only of the IFID and the number of interim file 1 positions, all blank. Doing this instrument by instrument should ensure that all files when processed and ready for battery merge will be of the same number of records with the same number of IFID's and that equal to what is on the SMIF file.

    A. Except for the fact that adults take a truncated battery. The specifications are given here for a single student master file development with adults and nonadults mixed and distinguishable only by IFID ranges. If separate adult files are separately developed, there is no problem; if the single file approach is used, it is not necessary to generate null records for SMIF nonmatches for instruments not given to adults. At battery merge time, legitimately blank fields can be placed on the master file, or variable length records can be used in the master file.

3. When generating null records, also place on interim file 1 a nonmatch code for all records so that scale scoring by programmed keys and range checking, can be bypassed in the type II program. Imputed values for missing scores are still desired for those who should have taken the instrument but did not.

4. In the case where completely blank answer sheets were received (except for dark-marked IFID), it will be necessary in the type II programs to program the bypassing operation on a test that all positions from the answer sheet(s) are blank. Actually, a few have stray marks, so that the test should be that 90% of the positions are blank.

5. If the matching operation between interim file 1 and SMIF file in step 1 above reveals the "theoretically impossible" situation where we have an instrument record but the student was not on the SMIF file, the student record should be deleted from the interim file 1 for that and all other instruments. It may be more sensible to accomplish this at battery merge time. Even though we may have complete battery information for such a student, we will not know his identification-descriptor set. The assumption here is that this will be rare (but may occur).

There should be a similar matching between the cover sheet file and the SMIF file, probably best handled by delivering an IFID tape pulled off from the SMIF file and sent to the link agency.

The ITED File

Three answer sheet sides per student must be scanned for the ITED scores, using the specifications for scanning facility operations. All subtests are power tests. In this and in subsequent tests, we assume that the scrambled file has been transmitted and unscrambled by a type I program. Starting with interim file 1, we consider the detailed specifications for the scoring and preliminary editing of operations in program type II, and note any further operations required to carry the file to the battery merge stage.

The scoring function of the type II program will develop 6 raw scores (RC, VOC, LU, SP, Math, and use of sources) by direct programming of the "rights only" scoring keys. There will probably be a separate key for each score, and each score will presumably come from (part of) one side of one answer sheet.

In addition to these raw scores, three are generated by direct summing: Reading Total = RC + Voc, Language Arts Total = Lang. Use + Spelling, and Composite Score = Reading Total + Language Arts Total + Mathematics.

Additional raw scores for Social Studies and Science need special treatment because each consists of a basic "background" component and a component which overlaps a subset of the Reading Comprehension subtest. MISOE requires both components, separately and combined into the total score, for each of the two content areas. The assumption is made here that the Social Studies Total score and the Science Total score are each provided by "total" keys and that component subscores from the Reading Comprehension key can be ascertained by comparing keys to define the overlap. These subscores should be programmed, and the nonoverlapping subscores obtained as generated variables by subtraction. (The programmed subscores can, of course, be obtained for the overlap subscores from Reading Comprehension generated by subtraction.) If, however, the independent parts of these scores are directly programmable from the keys, this should be done and the total scores computed as generated variables by addition.

The type II program should initially test for a null record and bypass to the write out. Otherwise the program should at this point result in generation of 15 raw scores: the six independent scores and three totals exclusive of the Social Studies and Science scores, the Social Studies Background score, the Social Studies component from Reading Comprehension, the Science Background score, the Science Background component from Reading Comprehension, and the total scores. Because these are developed from "rights only" keys, missing data from individual items will simply and properly result in a lower score. If raw scores are missing, imputation will be made by the type III program.

If conversion tables can be obtained from the test publisher, standard scores, percentiles, stanines, or "growth scores" will be ignored, but conversion to the IQ equivalent of the Composite Score obtained by programming the conversion table as a "table lookup" operation. With this, the ITED generates 16 scores on interim file 2.

No further metric change or generation of variables is required for the ITED interim file 2 created by the type II program. However, the range checking operation is required. The appropriate ranges for the 15 raw scores can be defined as zero to the number of holes in the appropriate keys. For the IQ equivalent, if generated, the range is defined by the IQ equivalent of the range on the raw Composite Score. For this set of tests, out-of-range scores will be replaced by blanks and the values imputed by program type III. Caution: when generating the three composite raw scores by adding key-scored components, program to check that the key-scored components are all present in the record; if not leave the field for the composite score blank, and in the case of the Composite Score, do not convert to IQ (leave IQ field blank).

The resulting interim file 3 should now be run through the frequency distribution program, printing out the frequencies of all scores, including zeroes and blanks, separately, and with separate distributions for the secondary OE, non-OE, and postsecondary students. The fields from ITED for adults will be legitimately blank. These distributions should be examined for reasonableness. These distributions can be moderately grouped in the tails but should be more finely broken out near the middle of the range. The printout should show both basic and cumulated frequencies and/or percents, so that the best imputation values can be chosen (probably medians in this case since there may be multimodality from too fine a division of scale).

Program type III will read in interim file 2 and impute the chosen values to blanks, outputing file 3 for merge and IFID-PID conversion. Blanks in ITED fields for adult records are legitimate and will, of course, not have any imputed values, and will remain blank.

The DAT File

Only three of the Differential Aptitude Tests are included in the student input battery: Clerical Speed and Accuracy, Mechanical Reasoning, and Space Relations. All three are administered to adults as well as to secondary and postsecondary students. Because the Clerical Speed and Accuracy test is a speeded test and the other two are factorially sensitive to time limits, the publisher's recommended time limits should be strictly observed. In the case of the Clerical Speed and Accuracy test, there are two parts: Part I for practice is not scored; Part II is scored and the publisher recommends that the student have two pencils since breakage and replacement will lose time and lower the score. (It is probably not a bad idea to have two pencils per student throughout all of the testing to minimize disruption and distraction in the testing room.)

The Digitek type of scannable answer sheets available from the publisher are designed as one double answer sheet, i.e., two perforation-attached answer sheets with both sides used, and for the entire DAT, in Forms L or M. Moreover, MISOE plans to use Form A of the Clerical Speed and Accuracy Test. For scanning purposes, the following things must be ascertained and treated accordingly:

1. Exactly how many answer sheets and sides are involved?
2. Whether each answer sheet that is separately scanned, after deperforation if a double answer sheet is involved, contains the IFID number. If not, special care must be taken to ensure that data from an individual student is properly collated through the processing.
3. Whether unused portions of the commercial answer sheets must be scanned, yielding unnecessary blank fields on the scrambled output tape, or whether the scanner can be simply programmed to avoid this.

Special control forms prepared in the scanning facility, if accurate, may be the simplest answer. The missing data select option may also be helpful.

4. Whether any of the answer sheet sides can be ignored.
5. Whether the best solution to these several difficulties is the preparation, with publisher permission, of special scannable answer sheets and control forms by MISOE.

There should be fields of ones and blanks on unscrambled tape for the three subtests given. Each of the three may be simply scored by a single "rights only" key programmed in the type II program. No metric changes, generated variables or a priori imputations are contemplated. Range checks should be made using ranges defined by zero (distinguished from blank) through the number of holes on the scoring key. Distributions should be made separately for secondary, postsecondary, and adult students and medians imputed by program type III. (It is assumed that these groups can be defined by IFID number groups, since the common descriptors are not on these interim files.) The distributions should also be checked for reasonableness.

The SSHA File

Only one side of one answer sheet is required for the Survey of Study Habits and Attitudes. Assuming that publisher permission has been granted to prepare a Digitek type answer sheet and control (or that these are available, despite no mention in the SSHA manual) and assuming its reading results in ones or blanks for each item, the type II program must score initially for 8 scores because each of the four basic scores is the sum of two components, one from a "rights" key and one from an "eliminator" key (a reversed "wrongs" key in this case). Stencil 10 for use with IBM 1230 answer sheets in coordination with the test form should define the four "rights" components: DA, WM, TA, and EA. It may be necessary to procure an information copy of the IBM 1230 answer sheet to ensure the correct coordination. Stencil 20 for use with IBM 1230 answer sheets should define the "eliminator" components for each of the four scales. Program internally to add the rights and eliminator scores for each of the four scales to produce the four R+E raw scores. Then program to add DA+WM to define the raw score for SH; TA+EA for the raw score for SA; and SH+SA for the raw score SO. Output these six scores. No metric changes, further generated variables, or a priori imputations are contemplated. Caution: in program type II, check that the response fields for items defining each of the four basic scores are not all blanks; if they are, bypass the scoring operations for that basic score so that it will be blank and not 00. Similarly, if any of these basic scores are blank, bypass the additions that yield the higher level scores. Range check basic scores 0-50, SH and SA 0-100, and SO, 0-200.

Distributions should be obtained for secondary and postsecondary students (adults will not receive SSHA) and median values imputed for the blanks in the seven fields (even for the higher level scores where the medians are not the sums of the medians on their components). SSHA fields for adults are left legitimately blank. Again, check the distributions for reasonableness.

The Personal Values File

The Survey of Personal Values (SPV) and the Survey of Interpersonal Values (SIPV) are very similar instruments and because of a common answer sheet problem, it will be convenient to treat them as a single instrument with two subtests. Since the Digitek answer sheet is not commercially available from the publisher, MISOE-designed sheets are required, but the test-item format as published and the OPSCAN-100 grid do not align. Therefore, it was decided to try to design a special form of the tests, with publisher permission, in which the tests and the answer sheet arrangement would be compatable. It was judged that the two tests could be administered together, using three answer sheet sides, i.e., approximately 1 - 1/2 sides or less per test. The following specifications are presented on the basis of that assumption. Each test has 30 items in triad formation, each member of a triad having a "most" and a "least" response alternative. Thus, there are 30x3x2=180 response positions for each test or 360 for the combined test. It is assumed that the interim file 1 record will, in addition to the IFID, have 360 positions with ones or blanks. Each test yields six raw scores ("rights only"), or 12 altogether. It is very important that the commercial scoring keys be perfectly coordinated with the item triads, which are unnumbered in the original test, and moreover, are split into two groups (A and B) of 15 triads within each test. If this is done, the basic scoring in the type 2 program, yielding two sets of six 2-digit scores in range of 00-32 each should prove no more difficult than the scoring of the other tests.

When the two sets of six scores have been obtained and held in storage by the program, they should each be summed to generate two validity check variables. If either of these two variables lies outside the range 85-95, replace the six score positions that generated it with blanks. If any of the 12 basic scores are out of the 00-32 range, replace it with blanks. No metric changes, other generated variables, or a priori imputations are contemplated. When outputing the records on the interim file 2, output only the IFID and the 12 basic scores; do not output the validity check variables. In the type II program, printout the number of validity check failures for both subtests. If this is unreasonably high, it would be prudent to try to ascertain the reason before proceeding with operations on the interim 2 output file.

Distributions should be run by secondary, postsecondary, and adult groups and medians chosen for imputation of missing values in the 12 score positions.

The HSPQ File

The instrument is available with a single Digitek answer sheet (one side only) from the publisher. Thus no special problems are anticipated for scanning operations, so long as the answer sheet form has unit-position defined fields, yielding the set of ones and blanks as presumed in the other tests. The scoring operations in the type II program generate 14 2-digit primary factor scores directly by weighted scoring. The scoring key (there may be two of them, each for a subset of the factors) has printed on it the weights "1" or "2" (no hole if weight is zero). Thus, the score is the sum of the weights for "holed" items: if the position is nonblank, program to add 1 to the factor score for a hole marked "1", and to add 2 to the factor score for a hole marked "2". One can program this (and other key-scoring for multiple scores) either by testing each position for additions to all scores or by rechecking the set of positions for additions to each score in turn; both are algebraically correct and the choice a matter of programming efficiency and convenience. Where scores are completely independent, it will usually be more convenient to take the items in turn.

No a priori imputations or generated variables are contemplated. It was decided to use only the 14 primary factors with no attempt to generate or use the second order factors. Although a number of metric conversions are provided for this instrument, none appear to be especially relevant to anticipated MISOE usage. Percentiles and other metric conversions based on MISOE groups can be obtained in the analysis system. The range checking operation should be included in the type II program with ranges from 00 to the maximum for each given scale. This maximum may be found by computing the weighted score for a hypothetical individual who marked all of the item positions for which there is a key hole for that factor.

Distributions should be run for secondary and postsecondary groups (test not given to adults). Medians should be inputed for missing factor scores. It is not likely that there will be many except perhaps for those who fail to complete the test (taken ill, etc.). Distributions should be run anyway for reasonableness checks.

The Culture Fair File

This instrument is also assumed to be available with a Digitek answer sheet (probably a single side). It yields four basic raw scores of two-digits each from "rights only" keys. On one of the keys, double responses to item alternatives are required so that two positions must be examined and addition to the raw score made only if both responses are made. Otherwise the scanning and the type II program operations appear to be straight-forward. A fifth variable is generated as the sum of the four scores; if any of the four is missing, leave the fifth variable score as blank,for distribution imputation by the type III program.

This instrument purports to be an intelligence test and an IQ conversion table is available. This should be included in the type II program and the IQ computed from the generated sum score by "table lookup" as a metric-changed variable. Both the raw sum score and IQ will be retained.

Distributions of all six scores should be made for checking and median location for secondary and postsecondary groups (no adults being given the test). Median values will be imputed for blanks by the type III program.

The MPI File

The extensive personal information questionnaire exists in separate forms for the nonadults (MPI) and for the adults (MALPI). They are very similar in content and format and may be discussed together. However, they are sufficiently unique, that they probably should be separately processed. The exact number of answer sheets has not yet been defined, but it is assumed that an interim file 1 with IFID match code has been produced. Specifications are therefore given in this section for the coding and editing operations in the type II program and the type of distributional values to impute to missing data by the type III program. The coding in effect includes variable generation and metric definition.

These specifications are presented systematically in tabular form, which should facilitate the computer programming. Specifications for MPI processing are given in Table 1; those for the MALPI in Table 2. The form of the tables is identical. In column 1 is the number of the item as it appears in the inventory; an asterisk by the item number means that there is a special matter to be dealt with in the text below. Column 2 shows the number of variables, all of which are one digit and therefore require one tape position. The total number of variables should give the record length when added to the 5 IFID positions (ignoring a match code position).

Column 3 specifies the coding of the variables. For those items where the response alternatives form ordered categories, the range of codes is given, n-m assigning to the first alternative (A), with n greater than m when coding from high to low and m greater than n when coding from low to high (the more usual case). For items in which alternatives are unordered or only partially ordered, dichotomous coding of each response alternative is indicated by 1/0 and means that a "1" is retained and a blank is replaced by zero. In some cases with partially ordered categories, those that are ordered receive the usual treatment and the unordered category which is not part of the scale is left blank for modal imputation by the type III program.

Column 4 indicates the treatment of blanks (missing data) and column 5 the treatment of multiple response. Typically, the dichotomously coded items receive "0" imputation for blanks, but an occasional exception occurs. In the event of multiple response with such items, no action is required (indicated by

Table 1

Coding Specifications for the Items in the Massachusetts Pupil Inventory

Item No.	No. of Variables	Codes	Blanks	Multiple Responses
1	1	1-5	3	3
2	5	1/0	0	NA
3	2	1-3	1	1
4	1	3-1	Mode	BM
5	1	0-5	0	BM
6	1	3-0	0	BM
7*	1	1/0	0	1
8*	1	1/0	0	1
9*	9	1/0	0	NA
10*	2	1/0	0	NA
11	5	1/0	0	NA
12	4	0-3	Mode	BM
13	1	0-5	Mode	BM
14	1	1/0	0	0
15	1	0-4	Mode	Blanks to mode
16	1	1-4	Mode	BM
17	1	1-6	Mode	BM
18*	--	--	--	--
19	1	1/0	1	1
20	1	1-5	Mode	BM
21	1	1-8	Median	B/coded median
22	1	1-4	Mode	BM
23	1	0-4	Mode	BM
24	1	0-5	Mode	BM
25	6	1/0	0	Replace with 0
26	1	1-9	Mode	BM
27	10	1/0	0	NA
28	1	0-4	Mode	BM
29	1	4-0	Mode	BM
30	1	1/0	0	0
31*	1	0-6	0	HV
32*	1	0-4	0	HV
33*	3	1-4	NA	Blank
34*	1	0-4	0	HV
35*	3	1-4	NA	Blank
36	1	0-3	0	2
37*	1	2,3,1	NA	
38	2	1-3	2	2
39	2	2-0	1	1
40	1	5-1	Mode	BM
41	7	1/9	9	NA
42	1	0-6	Mode	BM
43	1	0-5	Mode	BM
44	1	0-5	Mode	BM
45	1	3-1	Mode	BM
46	9	1/0	0	NA

(Continued on next page)

Table 1 (Continued)

Coding Specifications for the Items in the Massachusetts Pupil Inventory

Item No.	No. of Variables	Codes	Blanks	Multiple Responses
47	9	1/0	0	NA
48	1	5-1	Mode	BM
49	1	2-0	Mode	BM
50	2x6	1/0	0	NA
51	2	1-9	Mode	BM
52	4	1/0	0	NA
53	5	1/0	0	NA
54*	1	3,1,2	2	2
55	2	4-1	Mode	BM
56	2x3	1/0	0	NA
57	2	4-1	Mode	BM
58	2	5-1	Mode	BM
59	2	5-1	Mode	BM
60*	2	4-0	Mode	BM
61	2	4-0	Mode	BM
62	3x5	1/0	0	NA
63	2x8	1/0	0	NA
64	1	4-0	Mode	BM
65*	5	1/0	0	NA
66-77	12	3-1	2	2

NA) because the dichotomous coding takes care of it and it is usually legitimate anyway (e.g., mark all that apply). With ordered category coding, the usual situation is for blanks to be left blank in program II for imputation of the code for the modal category (occasionally the code for the category containing the median) by the type III program. In such items, multiple responses are replaced by blanks in the type II program to receive modal category codes by imputation in the type III program (indicated by BM in column 4). When values other than zero, mode , or median are indicated in columns 4 or 5, these are a priori values for programmed imputation in the type II program. It remains to deal with special problems involving the processing of those items indicated by an asterisk in column 1. We do this first for the items from Table 1 for the MPI.

Item 7: If marked and coded "1", force blanks into all record positions for responses to questions 8 and 9 before editing them. If "0", proceed.

Item 8: If a "1", force blanks into all record positions for response to question 9, before editing them. If "0", proceed.

Item 9: As worded, multiple responses to the question are illegitimate, but with dichotomous coding and item content, multiple responses are not meaningless. Moreover, no basis for a priori imputation exists and modal imputation involves some special problems (comparison of nine dichotomous distributions, etc.)

Item 10: Students can branch to item 10 from either a "yes" to item 7, or a "no" to item 7 and a "yes" to item 8. However, they are instructed to answer item 10 only in the former case, the direction being given before item 10 and therefore possibly missed. The recommended coding in item 10 is maximally flexible for any normal resolution of this situation.

Item 18: This item presents several problems. First,the booklet directions tell the student to write in the choice number in a single response position. It is likely that the scannable answer sheet will have 112x3 response positions so that the student may leave all three blank if not a choice, mark the first position for first choice, second position for second choice, and third for third. With this arrangement, one could code three 3-digit variables with codes from 001 for Accountant to 113 for Other and 000 for nonresponse or multiple choices at a given level. While this arrangement preserves all the data and permits logical data processing operations, it is not suitable for regression analysis which should use dichotomies or choice level codes; such would be awkward (or at least inefficient) to generate ad hoc in the analysis system. Therefore, despite the much larger number of variables (all of 1 digit), it is recommended that 113 variables, one from each type of work be generated using the following code:

0 if not chosen (positions blank for all three choices)

1 if third choice

2 if second choice

3 if first choice.

If a work type is chosen at more than one level (multiple response of one kind), code the sum of the codes. Thus the range is 0-6. The other kind of multiple response, theoretically illegitimate, where more than one type of work receives the same choice level(s), poses two alternatives. In one, we consider the illegitimacy overriding and code 0 in all such work types. This is not recommended because it requires the program to scann the response pattern by level over all 113x3 positions; possible, but not worth it. The other, which is recommended, is to follow the above coding scheme even for this type of multiple response; this is simpler and may be reflecting real ties in the students choices.

Item 31: If the code is "0" for "none" or by imputation for nonresponse (blank), force zeroes as codes for the variables generated from questions 32 and 34, and force blanks for the variables generated from questions 33 and 35. If multiple response is given to item 31, code the higher value (HV).

Items 32 and 34: Take the higher coded value for multiple response unless already destroyed by the forced zero rule from item 31.

Items 33 and 35: Blanks either from nonresponse or forced from editing operations in item 31 are legitimate and should be left that way. Multiple responses should not be coded but replaced by blanks.

Item 36: If "0" from "none" or replacement of blanks, force a blank in the variable from item 37.

Items 37 and 54: Note that the order of response alternatives is not the same as the ordering of the codes.

Item 60: An item in this form in which all categories but the last, usually "don't remember" or "don't know", are scalable, has a number of coding options. The compromise specified here retains analytic flexibility and yet permits scaling by those who do remember and respond. This is one of two examples in MPI; there are many in MALPI which, except for the addition of the "don't remember" category are comparable with an item in MPI. The recommended procedure, which also permits maximum comparison of remembering adults with nonadults, is:

a. dichotomously code each response alternative as shown in the tables

b. generate 3 scaled variables, coded in this case 4-1 for the regular categories A-D, leaving the scaled variables blank for those checking E or giving multiple responses, or having all zeroes after performing step a.

c. impute the modal (or a state a priori) value for those blanks in the generated variable in the type III program. For this item, impute modes.

Item 65: Add a generated variable coded 0-3 for A-D and impute the modal code.

The similar notes for items with an asterisk in Table 2 for the MALPI follow:

Item 8: Force zeroes into the 9 dichotomies in item 9 for those who are "1" in the first variable.

Item 11: Typographic error; two alternatives labeled C. Alternatives therefore go from A to G or seven.

Item 12: See discussion of item 18 for the MPI.

Item 13: Categories NO and YES have been reversed from that for the comparable item in MPI, which is more normal format. Specification of dichotomy is therefore reversed in Table 2.

Item 22: This item appears to be very faulty in construction; perhaps it is a typographic error in which additional categories have been omitted. Therefore, no specifications are provided at the present time.

Items 26-30: See notes for corresponding items 31-35 of the MPI; they are applicable if "5" is added to the MALPI item numbers to get the MPI numbers discussed previously.

Item 31: It should be noted that those responding "none" are told to skip four questions, where the analogous item 36 in the MPI tells the student to skip only one question. This should be checked out. If correct, force blanks into positions for items 32-35 of MALPI if the first variable ("none") is "1". Also, this is the first of several MALPI items analogous to items in MPI bit with a "don't remember" category added. Therefore, the noted coding retains dichotomies with the addition note here to generate the scaled variable. In this case, it is 0-4 for responses to A-D, blank for E or those all zeroes in A-D, to be imputed by the modal response code.

Item 32: A "Don't Remember" addition; dichotomous coding with an additional generated variable coded 2,3,1 for A,B,C (recall analogous MPI item 37) with"2" as the imputed value.

Table 2

Coding Specifications for the Items in the Massachusetts Adult Pupil Inventory

Item No.	No. of Variables	Codes	Blanks	Multiple Responses
1	1	0-9	Mode	BM
2	1	1-5	3	3
3	6	1/0	0	NA
4	2	1-3	1	1
5	1	3-1	M	
6	1	0-5	0	BM
7	1	3-0	0	BM
8*	4	1/0	0	NA
9	9	1/0	0	NA
10	6	1/0	0	NA
11*	7	1/0	0	NA
12*	--	--	--	--
13*	1	0/1	0	0
14	1	1-5	Mode	BM
15	1	1-8	Median	B/coded median
16	1	1-4	Mode	BM
17	1	0-4	Mode	BM
18	1	0-5	Mode	BM
19	6	1/0	0	Replace with zeroes
20	1	1-9	Mode	BM
21	10	1/0	0	NA
22*				
23	1	0-4	Mode	BM
24	1	4-0	Mode	BM
25	1	1/0	0	0
26*	1	0-6	0	HV
27*	1	0-4	0	HV
28*	3	1-4	NA	Blank
29*	1	0-4	0	HV
30*	3	1-4	NA	Blank
31*	5	1/0	0	NA
32*	4	1/0	0	NA
33	2	1-3	2	2
34*	2x4	1/0	0	NA
35	1	5-1	Mode	BM
36*	4	1/0	0	NA
37	9	1/0	0	NA
38	9	1/0	0	NA
39*	6	1/0	0	NA
40*	4	1/0	0	NA
41*	15+16	1/0	0	NA
42	2	1-9 for A+I Blank for J	Mode	BM
43*	6	1/0	0	NA
44*	4	1/0	0	NA
45	2x3	1/0	0	NA
46	2	4-1	Mode	BM

(Continued on next page)

Table 2 (Continued)

Coding Specifications for the Items in the Massachusetts Adult Pupil Inventory

Item No.	No. of Variables	Codes	Blanks	Multiple Responses
47	2	5-1	Mode	BM
48	2	5-1	Mode	BM
49	2	4-0	Mode	BM
50	2	4-0	Mode	BM
51*	3x6	1/0	0	NA
52	2x9	1/0	0	NA
53*	6	1/0	0	NA
54*	5	1/0	0	NA
55-66	12	3-1	2	2

Item 34: Add two generated variables coded 2-0 for A-C,with "1" the imputed value, one from dichotomies for "you" and one from those from "friends".

Item 36: Add a generated variable coded 3-1 for A-C and impute modal code.

Item 39: Add a generated variable 5-1 for A-E and impute mode.

Item 40: Add a generated variable 2-0 for A-C and impute mode.

Item 41: Note that no response position is available for Father as housewife; this is correct and leads to 15 positions for father and 16 for mother. Note however, this is not consistent with MPI item 50 which should be adjusted accordingly in both answer sheet and Table 1 specification.

Item 43: Leave as dichotomies only; no generated variable.

Item 44: Add a generated variable coded 3,1,2 for A, B, C with imputed value of "2".

Item 51: Add three generated variables coded 4-1 for A-D with modes for imputed values.

Item 53: Add a generated variable coded 4-0 for A-E with modal codes imputed.

Item 54: Add a generated variable coded 0-3 for A-D and impute the modal code.

V. Data Entry Operations With the Student Process Battery

The process portion of the student master file will be developed from administration of the School Sentiment Index and the Student Program Questionnaire to students while in their programs, and the completion, for each of their students, of the Master Identification Form Update (MIFU) by department heads. The battery is to be replicated over time so that the processing cycle from administration to file merge will be repeated a variable number of times depending on program length. Each instrument on each replication will generate an interim file 1 to be processed.

The last replication cycle for the battery occurs at end-of-program time. This is also the time at which process data are to be collected, some from students, some from department heads, and in the case of ratings on terminal objectives, an additional data source group. It is anticipated that certain aspects of the logistics of data collection and processing in product space may overlap in some sense that of the final cycle of process space development (within a program), certain modifications of the final cycle of process may be indicated. These issues will be discussed in more detail in the next chapter.

A new cover sheet, designed and processed like that for the student input battery, but labeled as the cover sheet for the process battery is required for each over time replication. The IFID numbers associated with the input battery can no longer be used to dark-mark the answer sheets for the process battery because there is no way to ensure that the answer sheets with a given number will get completed by the original student. Even if the schools have IFID-name rosters, maintaining the correlation between students, IFID numbers, and answer sheets would be a logistic hazard.

The cover sheet for the process battery will define that correlation as before, but the link agency will have to make a name, address, and date-of-birth match to develop the new IFID link to the same PID so that process data can be merged to the student input data. This arrangement would take care of the coordination of the two instruments to be completed by the students and ensure mergability to the master file. However, the MIFU forms to be completed by department heads would have only coded information so he would not know which student's attendance information to look up and code. One solution to this is for the MIFU to be part of the packet going to the student and immediately following the cover sheet. On completion of the cover sheet, the student would then remove both the cover sheet and the MIFU form, which must be attached to the cover sheet as a perforated foldout to keep them temporarily together. When this combination is placed in a receptacle for delivery to the link agency, it first goes to the department heads, who complete the MIFU forms while the cover sheets with names to look up are still attached. The department head then separates the two forms, sending the cover sheet to the link agency and the MIFU (with the completed student process tests, or separately) to the scanning facility. Note that there is no information on the cover sheet or the MIFU which is not known or knowable to the department heads and hence no violation of confidentiality is involved. It will, however, be necessary to "level with" the students about the need and use of MIFU forms, even though they do not fill them out. There is a further detail in the testing session logistics with this: instead of students placing the combination of cover sheet and MIFU in a single box or envelope as in the student battery, they would place them in separate departmental piles, unless testing sessions can be arranged on a departmental basis.

We proceed now to discuss the processing specifications for the interim files by type II and III programs.

The School Sentiment Index Files

This instrument consisting of 83 four-choice items should be simple to set up for the optical scanner, yielding 332 tape positions per record plus the new IFID from dark-mark sensing. The four positions per item on the interim file 1 should first be changed in the type II program to one position per item with the response codes, and then six variables generated. Each item score as coded contributes to one of five generated factor variables, scores on which will be added to compute the sixth or Total Score. The five factors are:

T: Attitude toward teachers and teaching (39 items)

S: Attitude toward the structure and climate of the school (20 items)

P: Attitude toward peers and peer relations (6 items)

L: Attitude toward learning (7 items)

G: A general summary of residual attitudes, feelings, and behaviors (11 items).

The items are coded 1-4 or 4-1 depending on whether statements are worded or interpreted (in terms of student perceptions) positively or negatively. Thus, the specifications consist of stating for each item the factor to which it contributes and which way the codes run. This information is given in Table 3 with the item number in column 1, the factor initial in column 2, and the code for the "Strongly Agree" response category. Thus, item 1 is coded 1-4 and contributes to the T-factor, while item 2 is coded 4-1 and contributes to the G-factor.

With the absence of a middle, "indifferent", category, the a priori imputation for missing data, required before factor scores are computed, presents a minor problem. Taking into account response set considerations, impute the code for "disagree" for the 4-1 coded items and the code for "agree" for the 1-4 coded items; conveniently, this turns out to be "2" in both cases. Therefore, impute a "2" for missing values or multiple responses, when coding and before computing the factor scores. Each factor score is the sum of the item scores keyed to that factor as shown in Table 3. When the five factor scores for a record have been computed, they are to be added to get the sixth, or Total Score. All of this is to be accomplished in the type II program with interim file 2 containing the header label and each record containing the new IFID, the 83 item scores, and the six generated scores. The usual range-checking operation should also be included in the type II program.

Distributions of the six factor scores should be run for a reasonableness check. If the rules followed for the student input battery are followed here with regard to records for students who missed the instrument, there should

Table 3

Coding Specifications for the School Sentiment Index

Item	Factor	Code[a]	Item	Factor	Code	Item	Factor	Code	Item	Factor	Code
1	T	1	22	G	1	43	T	1	64	S	4
2	G	4	23	S	1	44	T	4	65	T	4
3	T	4	24	S	1	45	T	1	66	P	1
4	G	4	25	T	1	46	S	4	67	T	4
5	S	1	26	P	4	47	T	4	68	L	4
6	T	4	27	T	1	48	T	1	69	T	1
7	S	1	28	L	4	49	S	4	70	G	1
8	T	1	29	T	1	50	T	1	71	S	4
9	P	1	30	G	1	51	P	1	72	S	4
10	T	4	31	S	1	52	T	4	73	T	4
11	L	1	32	S	4	53	L	4	74	T	1
12	T	4	33	T	4	54	G	4	75	L	4
13	G	1	34	T	1	55	G	4	76	T	1
14	S	1	35	L	4	56	S	1	77	T	4
15	T	4	36	S	1	57	T	4	78	S	1
16	S	4	37	S	4	58	T	4	79	T	1
17	T	1	38	T	4	59	G	4	80	G	1
18	P	4	39	T	4	60	T	4	81	T	1
19	T	4	40	S	1	51	G	4	82	T	1
20	L	4	41	T	1	62	S	1	83	S	4
21	T	4	42	P	4	63	T	4	--	--	--

[a] The code for "Strongly Agree". See text.

be some null records on interim file 2. Impute "2" for the item scores and medians for the factors scores in the type III program. Interim file 3 must be merged with those from the MIFU and the Student Program Questionnaire (as described below) from a given replication cycle, matching on the new set of IFID numbers; the merged file must then have PID replacement by a process-cycle scramble file from the link agency.

It should be noted that one difference obtains in the rules for handling the case of records for students missing an answer sheet and that is that the matching must be done against the MIFU rather than the SMIF file with its different set of IFID numbers.

The Student Program Questionnaire Files

Although in semantic differential format, this 20-item instrument aimed at tapping the student's attitude toward the program in which enrolled, will be treated quite simply in MISOE. The 20 items with their seven response alternatives will be coded as 20 item scores on a 1-7 or 7-1 scale with "3" imputed a priori for any unanswered or multiple-answered item, and a Total Score generated as the simple sum of the 20 item scores. It remains to define the coding direction for each of the 20 items; this follows in the form of stating the left-side word, followed by a "1" or "7" indicating the left-side "very" response alternative:

Worthy - 7
Unsuccessful - 1
Interesting - 7
Satisfactory - 7
Unrewarding - 1
Impractical - 1
Desirable - 7
Unessential - 1
Effective - 7
Important - 7

Harmful - 1
Worthless - 1
Meaningful - 7
Unrealistic - 1
Definite - 7
Attractive - 7
Profitable - 7
Aimless - 1
Insecure - 1
Disreputable - 1

Check the range from 20-140. Distributions of the Total Score should be obtained for inspection and median value (presumably about 60) imputed.

The Master Identification Form Update Files

The MIFU instrument serves a function quite analagous to the SMIF in the input battery, but in addition serves two other functions. One is to update the grade and program descriptor information, and the other to obtain attendance information. The specialized relation to the cover sheet has already been discussed, as has the analagous use of this file to detect missing answer sheets (SSI and SPQ) and generate null records. It is therefore like the SMIF in its requirement that it be completely filled out. Also like the SMIF, it is an exception to the generation of a tape file of only "ones" and blanks, updated codes and attendance figures being scan-coded.

The processing for this file would appear to be quite analagous to that for the SMIF, except that information from the MIFU is placed in the process section of the student master file at merge. It is tentatively assumed that attendance and other information sought from department heads can be completely provided; if not, median numbers of days present or tardy, etc. would seem to be reasonable imputations. It should be possible to determine reasonable ranges for these variables.

On the last replication of the battery, the MIFU may be extended to permit department heads to add codes for completion/noncompletion and information on noncompletors.

VI. Data Entry Specifications for the Student Product Battery

The product space battery consists of: (1) whether the student did or did not complete the program and special information on the noncompletors, (2) ITED retest for all secondary (OE and non-OE) students in SDS-2 and postsecondary students expected to complete an associate degree; thus, certificate level postsecondary students are not retested on ITED nor adults retested on DAT; and (3) ratings on terminal objectives.

The information on program completion and special information on noncompletors are most readily picked up on an extended MIFU administered with the last cycle of the process battery. Although the MIFU is usually filled out by the students, recall that it was recommended that it remain attached to the cover sheet so that department heads could code the enrollment and attendance data and then be separated, with the cover sheet going to the link agency and MIFU going to the scanning facility. With this same arrangement, the department head can supply on the extended MIFU the completion data. For completors, this will consist only of the fact of completion, with the fields for information on noncompletors being left blank; indeed, blanks should be forced onto the tape files for the corresponding fields.

For those receiving the ITED retest, the set of ITED answer sheets can be added to those for the School Sentiment Index and Student Program Questionnaire when administering the last cycle of the process battery. These can then be returned to the scanning facility in the usual manner and processed in accordance with the specifications for the pretest in the input battery.

Before discussing the data entry problems for the ratings on terminal objectives, a few points need to be made concerning the integration of the above data collection with that of the last cycle of the process battery. One advantage is that we avoid one more round of new cover sheets, IFID numbers, and special name-matching by the link agency to ensure that all data is mergeable on a common set of PID numbers. It should be recalled that we already have in addition to the initial round with the input battery, as many rounds in process space as years of a program. The new IFID numbers for the last cycle of process testing will, of course, have to be dark-marked on the extended MIFU and retest ITED answer sheets and collated with the process battery answer sheets. Because the ITED retest is administered to a subset of students, the integration of its answer sheets with the rest of the final cycle process packet must be selective, as it was in the input battery, but the selection here is greater because only the associate level students take the retest in the post-secondary group. Blocks of IFID numbers must be preassigned to groups as in the input battery to control this and testing at the community colleges must have logistic separation of grade 13 and grade 14 students.

The main issue which remains is the data entry specifications for terminal objective ratings, starting with the logistics of collecting the information. Again, we would like to minimize the addition of one or more IFID-name-PID coordination cycles. The suggested logistics for collecting these data in SDS-2 programs follow:

1. Three answer sheets, one for each of a maximum, and hopefully a standard, number of raters per subject are included in the last cycle process battery packet.

2. The student is told to ignore these, except that they are removed as a subpacket with the combination of cover sheet and extended MIFU. The problem here is to keep these together and it may not be possible to do so by stringing out five answer sheets connected by performations. The problem appears to be one of differential binding or packaging for this subset of answer sheets.

3. The ITED answer sheets are collected and sent to the scanning facility in the usual manner.

4. The cover sheet, extended MIFU, and three terminal objectives answer sheets as a subpacket go first to the department head. If he is in charge of more than one program, as that is defined at that point in time, he

should receive the packets in subsets by different programs for which he is responsible.

5. Before he completes the product portion of the extended MIFU, the three raters of students in a given program are assembled and given the subpackets intact. These raters make their ratings from whatever stimuli they are given (televised, direct observation, etc.) with the department head helping to identify the individual student subpacket to be used.

6. On completion of the ratings, the raters place their completed answer sheets in a manila envelope prepared by MISOE and addressed to the scanning facility. Thus, the department head has minimum influence on the answer sheet marks or minimum opportunity for his further MISOE functions to be affected thereby.

7. The department head takes the cover sheet and extended MIFU back to his records for the completion/noncompletion information. He then separates the cover sheet from the extended MIFU, sending the former to the link agency, and the latter to the scanning facility.

The processing of the cover sheet has been discussed in connection with replicated process battery cover sheets. The general processing of MIFU forms has also been so discussed in the same connection; however, we now have additional information on completion and noncompletion. Thus, interim files will have additional fields. One position will be coded 1/0 for completion/noncompletion (noncompletors should be blank in this position on interim file 1 if a single position on the answer sheet is provided to be filled in for completors only, with the zero replacement occuring in the type II program). Assuming that the noncompletor data includes the date of exit, this can be coded in the same way the link agency codes the date of birth from the input battery cover sheet; if this information is in the form of number of weeks, months, and/or years in the program, it can be converted to the best common base for that program (better, a common base across programs, such as weeks). Reason for termination can probably be simply coded; the best approach is to use dichotomous codes on the reasons so that multiple reasons are readily handled. Theoretically, the department heads should be able to complete all forms for all students. Actually, there may be legitimate reasons why they cannot. Length of time of program (or date of termination) would probably require distribution information and imputation of median values within program by the type III editing program. Missing data on reasons are covered by a series of all zeroes across the reason dichotomies. Little range checking is required, except that length of time in program must not exceed the program length. No generated variables are anticipated.

In the case of the three answer sheets per student with their common IFID in the dark-mark portion of the answer sheet, two matters require attention. First, it might be desirable for rater reliability studies to have the raters identified. One way to do this is to use the sixth digit of the dark-mark field to precode 1,2, or 3. Then at rating time, the department head (or other responsible school official, so long as the cover sheets and extended MIFU forms get back to the department head) says to the raters, "you are #1, you are #2, and you are #3. Each of you are to rate that student over there (or this student, or Mr(s) Jones, or whatever, depending on the stimulus-observation technique used). Here is your rating form" (which has been previously explained to the rater). This rater number would go onto the interim files with the IFID until aggregation of mean ratings discussed below. It is suggested that the three answer sheets per student be read onto a single interim file 1, a seriatim. If there are two raters for any reason, the third (completely blank) answer sheet should not be processed. It is also conceivable that some students will have no ratings either because they did not complete the program, or were absent when the ratings were accomplished or were absent when their performance was teletaped. This is presumably the kind of situation previously discussed where a student misses a whole instrument, and null records need to be generated through matching with the MIF forms (MIFU in this case).

The second consideration requiring attention is the special interim file operations in the type II program, where the ratings need to be aggregated to mean values for each student on each objective. It is assumed that the objectives to be rated are laid out on the answer sheet in a systematic way, which is coordinated with the objective and level (units and blocks) being rated; and that this can be unscrambled in the normal way. We start then, with an unscrambled interim file 1 containing for each student the outcomes of zero to three answer sheets. The type II program must:

1. ascertain the number of ratings available, 0-3, for each individual objective. Call this n_{ij} where i is the index for the objective being rated for the jth student.

2. sum the n_{ij} ratings (which are nonblank), X_{ij}, across the same ith objective positions in the three rater or answer sheet macrofields.

3. divide the sum of X_{ij} by n_{ij} to get $\overline{X}_{ij}$ for each ith objective. (We are working on a single record for the jth student.)

4. Output on interim file 2 the header label, and for each student, the array of means on the terminal objectives. Some or all positions may be blank. Distribute the mean ratings for each objective within groups and programs and impute the median values in interim file 3.

Save the interim file 1 for in-house rater reliability analysis. It should be noted that interim file 1 records should carry both the rater code and the IFID from each answer sheet and that the type II program should check the adjacency and common IFID of rating fields for a given student. This can be ensured by reformatting the interim file 1 on rater number within IFID. In outputing the record from the type II program on interim file 2, the repetition of the IFID can be eliminated, being sure that the IFID is retained once for each student record.

In addition to the ratings on terminal objectives are three kinds of information about each objective. One identifies the objective by subgroup, block, and unit. This information is not needed in processing except to define tape layout and document positions for terminal objective information.

The second is the difficulty level, which can be coded 2-1-0 and should be on the tape for weighting the ratings in analysis, when and if desired. If these are preassigned they can be programmed into the file by the type II program. If they are separately rated by each rater, provision for this must be on the scannable answer sheet rating form and these difficulty ratings placed on the rating file with the corresponding performance ratings and the average difficulties placed on the product data file.

It should be noted that the average-weighted average performance rating is not the same as the average of the individual difficulty-weighted averages. If the latter is also desired, it must be computed in processing the interim file 1 as the sum of W_iX_{ij} divided by n_{ij} and the results posted to interim file 2. Whatever is decided here, the information for a given terminal objective should be brought and kept together as microfields within single objective macrofields on the data file layouts.

The third piece of information, whether the objective is psychomotor, cognotive, or affective, can be decided a priori and dichotomous codes programmed onto interim file 2 to permit ready selection of groups of objectives in analysis.

Finally, it should be recalled, that at merge time, the last-cycle process files and the product files, presumably now on a common IFID which is not the same as the input battery or earlier process cycle IFID's, are merged first on this common IFID, then converted by the common process-product scramble file to the PID that is common across IPP elements. When this has been accomplished, this mass of data can be merged onto the student master file.

VII. Data Entry Specifications for the Impact Batteries

The two impact batteries, cross-sectional and longitudinal, although essentially identical in content and most coding aspects, require quite different logistics of data collection. This follows from the facts that the cross-sectional samples have no input battery operations identifying individual names and addresses, and that different followup groups are defined by different cohort lags rather than by different followup times on a given cohort. Both, however, must maintain confidentiality requirements, and, in both cases, data collection is accomplished by mail contact with subjects no longer under logistic controls through the educational system. Also, in both cases, the information obtained from completed instruments must be partially hand-inspected, hand-coded, keypunched, and verified, rather than optically scanned. The common content of the two batteries consists of the Massachusetts Educational Impact Inventory, completed by former students, and the Massachusetts Job Evaluation Form, completed by the on-the-job supervisors of former students.

In a sense, much of the cross-sectional effort field tests the instruments for the longitudinal effort, in addition to its main purpose of obtaining some impact space data prior to maturation of longitudinal cohorts. Despite the laudable nature of these aims, their implementation contemporaneously with that of the initiation of the total MISOE, constitutes an enormous burden not only on the MISOE staff and the link agency, but also on the staffs of the LEA's in both SDS-1 and SDS-2. Even limiting the cross-sectional battery to SDS-2 would be little relief. Moreover, the dollar costs are far from negligible.

Given the greater lead time available for preparing the one-year longitudinal followups, with opportunity to concentrate on polishing logistics and instrumentation, some of the loss of the field-testing benefits of cross-sectional effort could be offset, and furthermore, information gained from the first followups in short programs can be used to further polish the system for those in the longer programs and for the longer range followups. Moreover, this approach does not preclude a genuine feasibility "field-test" on a sporadic sample of, say 200 former students, i.e., about 50 per cohort lag (with anticipated returns of about 25 per cohort lag). Admittedly, management would have to wait a while before having operationally usable impact data; but, that from the cross-sectional effort is so unconnected with anything as to virtually reduce analysis to item distributions and their statistics. Even if some group comparisons are made on such statistics, the conclusions to be drawn from them are somewhat hazardous. Given this weighing of the pros and cons, costs and benefits, it is strongly recommended that the cross-sectional effort be eliminated from further development, implementation, or analysis.

Longitudinal Logistics for the MEII

The following steps are suggested for "getting the ball rolling" for a longitudinal followup in impact space:

1. MISOE prepares a set of files containing PID numbers and selective identification-descriptor information from the student master files. Initially, a file is produced for the one-year followup of the shortest programs. Then, a file is produced for the next-length programs until all program lengths are covered (perhaps 3-6 files). It will not be necessary to repeat this operation for the longer range followups.

2. These files will go to the link agency for match with the student name and address files as updated from previous followups. It may be convenient, and reduce the total number of file matches in the link agency, to transfer

the names and addresses to the transmitted files (PID replaced by IFID), with these files being updated, rather than the original name and address files.

3. The link agency prepares mailing labels and mails out the MEII booklets, previously prepared in accordance with some specifications to follow, and transmitted to the link agency. The mailout packets must contain a return envelope addressed to MISOE. One possibly minor problem is that subjects receiving mail purporting to be MISOE business, but from a foreign country, may have suspicions aroused (they surely cannot be expected to remember the link system in any detail); one solution is to have a MISOE representative return the mailout packets to Massachusetts for stamping or metering and posting in the U.S. mail system.

U.S. [illegible]

4. The link agency sends a list of the IFID numbers or their ranges by program to MISOE for getting these on the MEII booklets. Perhaps these IFID numbers should be printed on every page of the booklet in case they are returned partially damaged. As a minimum, they should be printed on page 1. This would be less expensive and not require IFID number collation of booklet pages. The booklets must, however, be coordinated at the link agency with the names and addresses; perhaps the easy way to do this is to include the IFID numbers on the labels.

The returned forms must be carefully inspected and subjected to some degree of precoding prior to verified keypunching. When the keypunching has been accomplished, with the IFID numbers in the initial field of each card, and a card number on each card, serialized within the set, the first few cards containing name and address update information, (and only that information) must go to the link agency. Another subset of punched cards containing information about the supervisor and his company (and only that information) must be pulled out for special treatment. These separations must be made only after the sets of cards have been checked for completeness and ordering. The remaining subset containing the coded and punched data can be read to tape in order, thus producing an interim file 1. In order to minimize the amount of clerical coding, and to eliminate the hazardous coding-at-the-keypunch, it is recommended that the data be punched as a one column per response alternative with "ones" or blanks, except for dollar amounts and the necessary clerical coding of certain open-ended items. The rest of the coding and editing specifications can then be carried out using type II and III programs, as usual.

The subset of cards containing supervisor names and addresses, and also containing the IFID number of the student respondent, should be card-to-taped, and address labels prepared for MJEF mailout. Listings should be produced for dunning, checking returns, and nonresponding supervisors. It is assumed here that the already heavily burdened link agency need not be involved in these operations. When the MJEF data comes back, they can be processed in accordance with specifications given in a later section of this chapter and directly merged to the student's MEII data record in terms of IFID numbers of the students, IFID numbers replaced with PID numbers, and the whole merged onto the student master file.

If it is desired to keep the supervisor name and address files for any reason, they should probably go to the link agency after these operations are completed and IFID numbers replaced with PID numbers.

Coding and Editing Specifications for the MEII

When the MEII forms come in, they must be carefully inspected. First, are they intact booklets? If not, are they usable with perhaps an occasional page missing or the page otherwise not usable? In the latter case, processing can proceed. Did subject bog down and fail to complete, but nevertheless returned the booklet? Some judgement to process what is available vs. rejecting the return (and treating it as a non-dunnable nonresponse) should be made. If the volume of this problem is large, it may be possible from inspecting a sample to formulate more precise acceptance criteria.

The accepted forms must now have open-ended items hand-coded on the booklet and keypunching instructions make reference thereto. Which these are and how they are to be treated is discussed below.

The coding and editing specifications are given in Table 4. Items are identified in terms of their numbers on Form A(1-year followup) with cross-reference to corresponding item numbers on Forms B (for 3- and 5-year followups) and C (for the 10-year followup). Otherwise, the table is similar to earlier tables with columns specifying (with the same abbreviations) the number of variables, coding, and treatment of missing data and multiple responses. Also, asterisks are again used to refer to specific issues discussed in the ensuing text. Any item requiring preliminary hand-coding will be identified and discussed through this mechanism, in addition to specific issues for items requiring special attention in the type II or III programs. Otherwise, the specifications refer to type II and III program operations on interim files 1 and 2, as before. Note that items start renumbering within parts. The special notes for items in items Table 4 marked with an asterisk follow with the items identified in terms of the Form A number:

Part I

Item 2: alternative (e) in Form C should read "six or more"

Item 19: In form C, corresponding item is also numbered 19, but no item 18 is indicated; hence, table shows item as numbered 18.

Part II

Item 1: If NO (blank), force blanks into response fields corresponding to the intervening questions defined by the skip (2-10 in A, 2-7 in B, and not applicable in C). Keypunchers merely skip over the corresponding card columns. The type II program must test on this item and branch around replacement of blanks by zeroes.

Item 2: This item requires precoding by hand, based on name of job and things done on job, using the principles advocated by David N. Wheeler and the Dictionary of Occupational Titles. For each of one to five jobs (allow five macrofields) place code to the left and instruct the keypunchers to punch first the code and then the months held (allow 3 positions to cover the person who held the same job for 10 years). Coders should examine the job pattern, including response to unemployment item 2 in Part III in the light of the length of followup period to impute a "months held" value, when not given. If this does not resolve the matter, prorate the total months in the followup period across the number of jobs held. More space is needed for things done on the first (i.e., most recent) job. If less than five jobs were held, the remaining macrofields are blank. If respondent skips spaces, e.g., filled in the first, third, and fifth job fields for three jobs, coders should indicate to keypunchers to put the information in the first three macrofields and leave the last two blank, rather than have alternating filled in and blank macrofields.

Item 4: In addition to the 14 dichotomies, punched by the keypunchers as ones and blanks, the type II program should change blanks zeroes, unless the response to item 1 is "1", and generate four dichotomous variables for the four major employer types (ignore "other"). The coders should note externally any examples of write-ins under "Specify" for the "Other" for MISOE documentation. It is anticipated that this will be sufficiently rare and unimportant that no special codes are required on the file, beyond the dichotomy for "Other."

Table 4

Coding Specifications for the MEII

	Item Numbers			No. of Variables	Codes	Blanks	Multiple Responses
	Form A	Form B	Form C				
Part I:	1	1	1	6	1/0	0	NA
	2	2	2*	1	0-4	M	BM
	3	3	3	1	0-5	M	BM
	4	4	4	20	2-0	0	1
	5	5	5	8	2-0	0	1
	6	6	6	1	2-0	0	BM
	7	7	--	1	1/0	0	1
	8	8	--	1	1/0	0	0
	9	9	8	7	2-0	0	1
	10	10	9	6	1/0	0	NA
	11	11	10	5	1/0	0	NA
	12	12	11	1	1/0	0	0
	13	15	14	1	0-3	M	BM
	14	13	12	7	1/0	0	NA
	15	14	13	1	1-4	M	BM
	16	16	15	1	0-4	M	BM
	17	17	16	1	0-4	M	BM
	18	18	17	1	0-5	M	BM
	19	19	18*	11	4-1	M	BM
Part II:	1*	1	1	1	1/0	0	0
	2*	2	2	--	--	--	--
	3	3	3	12	4-1	M	BM
	4*	4	4	14+4GV	1/0	0	NA
	5	5	--	1	1-5	M	BM
	6	--	--	10	1/0	0	NA
	7	--	--	1	4-1	M	BM
	8*	6	6	--	--	--	--
	9	--	5	5	1/0	0	NA
	10*	7	7	--	--	--	--
	11*	8	8	9	amts	0	NA
	12*	9	9	24	amts	0	NA
	13*	10	10	73	amts	0	NA
Part III:	1*	1	1	6(5)	1/0	0	NA
	2*	2	2	1	1-4A,1-5B, 1-6C	0	BM
	3	3	3	1	2-0	0	1
	4	4	4	16	1/0	0	NA
	5	5	5	5	1/0	0	NA
Part IV:	1	1	1	1	5-1	M	BM
	2	2	2	9	4-1	M	BM
	3	3	3	16	4-1	M	BM
	4*	4	4	5/7/7/6 +3prog codes	1/0	0	NA
Part V:	1*	1	1	3	1/0	0	NA
	2*	2	2	7	1/0*	0	NA
	3*	3	3	1	1-3A,1-5B_3 1-6B_5+C	0	LV
	4*	4	4	1+code	1/0	0	1 if type given
	5*	5	5	1+2codes	1/0	0	1 if either code given
	6*	6	6	1	*	M	BM

This will be the general rule for these situations. Coders should also note whether the "Specify" write-in should actually come under one of the other 10 groupings. If so, coder should so indicate and cross-out the mark on "Other."

Item 8: Generous card space, even if it takes several cards, should be provided to ensure provided information is completely punched. Coders should supply missing ZIP codes from the ZIP directory. As previously noted, it is important that all punched cards carry the IFID numbers, and that this subset of cards after verified punching be read to a separate tape. Thus, this information will not be on interim file 1. However, the keypuncher should be instructed that the next data item, i.e., the first data item on the next card following the duplicated IFID, shall be a "1" if a supervisor was named and blank if not. This information will then go onto the interim file 1 for data processing controls and counts of those supplying supervisor information.

Item 10: The same principles for coding and punching this information indicated for item 2 apply here for the part-time jobs. If no part-time jobs are listed, blanks are forced into the five macrofields for this item. Note that the booklets do not provide a constant number of macrofields for full-time and part-time job listings across forms. This should be made constant at five.

Items 11 and 12: In item 11, allow 5 positions for (a), 4 each for (b) through (e), 5 each for (f) through (h) and 6 for the total. The 12 variables from item 12 include both the monthly total and the "times 12 total". Coders must examine these totals for accuracy of multiplication and agreement of the last total with that from item 11. If they do not agree, the respondent's explanation if any should be examined, as well as his additions in both items. Until some problem returns have been examined by MISOE staff, it will not be possible to state precise reconciliation rules to the coders in terms of adjusting or prorating part-whole relations to affect reconciliation. The AMTS for codes on these items and for 13 refer, of course, to the written-in amounts. In the case of item 11, coders should check that the "nearest hundred" rule applies consistently.

Item 13: In preparing these forms, the net worth was not defined (it was in the original) as the difference between market value and amount owed. Either this must be returned to the item stem, or better, omitted from the booklet for computation by the type II program. The latter has the advantage that we do not have to correct the respondents' subtraction errors and reduces the load on respondent. If this is done, there are 47 variables to be punched and 26 differences to be generated (for those X'd out under amount owed subtract zero).

Part III

Item 1: Form A has six alternatives, the others five; hence the different number of variables for the different forms. It might save confusion in future processing and analysis to keep the macrofield length constant. This can be done by having six variables for all forms with the fifth variable in forms B and C always zero, and the sixth variable representing the response to alternative (e) in those forms. If respondent was never employed, skip macrofields to Part IV.

Item 2: Again there are varying numbers of response alternatives across forms, and with different cutpoints. However, this is a single scaled variable rather than a set of dichotomies, so the matter is handled simply by the varying ranges on the codes as indicated in Table 4.

Part IV

Item 4 has varying numbers of response alternatives with dichotomous coding. The principle suggested for item 1 of Part III applies with seven positions allowed and all zeroes on the dummy positions for the other forms. Those checking any of the first three positions are asked to specify the program. Therefore, these must be coded in accordance with MISOE program coding rules (possibly the USOE codes will suffice) prior to keypunching. The card layouts must contain three fields for this information, any or all of which may be legitimately blank.

Part V

Item 1: If (a) is checked, or the whole item left blank, skip the card fields for items 1-6. Item 1 consists of three dichotomies from responses to (b), (c), and (d).

Item 2: Either put this in multiple choice format or have coders use the following code:

Army - 1

Navy - 2

Air Force - 3

Coast Guard - 4

Marines - 5

National Guard - 6

Other - 7

Item 3: This item has differing numbers of alternatives in different forms and should be treated like item 2 of Part III, in this respect. Note that the 3-year and 5-year Forms B are different on this item. Multiple responses should be punched with the code for the lower valued response. This item should be edited in the booklet so that exactly 3 or 6 months will not generate multiple responses.

Items 4 and 5: A single, not double dichotomy should be coded. The coders must code the type of specialist training and career aspiration names. Since the service codes for these vary with the branch of service, a coding scheme should be developed relating responses to DOT or other MISOE-consistent base.

Item 6: Since the rank names vary with service, a multiple choice format would be difficult even with a matrix form designed against the item 2 codes. To develop a coding scheme:

1. List the rank names from high to low rank for each service with cross-reference to common level ranks (e.g., ensign and second lieutenant) across services.
2. Assign common codes from high to low to the common ranks with skips for unique levels within service
3. Assign intermediate codes for unique ranks within service. This should provide a common coding base across the impact space for analyses involving military ranks. Leave nonresponse blank for imputation of the non-zero modal code by the type III program.

It is essential that precise instructions be developed for coders and keypunchers and that their operations be carefully monitored to ensure maximum quality control of the interim file 1 data.

Longitudinal Logistics for the MJEF

Some of the data collection specifications for MJEF have been given earlier in this chapter. It is important that the student's IFID number be printed on the first page of the form and that the mailout packet contain a return envelope addressed to MISOE. The same quality control rules for coders and keypunchers specified for MEII apply to MJEF.

Specifications for this single form are given in Table 5, beginning with item 3. The first two items must be examined and coded in terms of the David Wheeler and DOT rules used in similar situations for MEII.

In items 4-6, keypunchers punch number of years and number of months as separate fields leaving them blank if they are blank. The type II program should read these paired fields and converted each pair to the number of months basis = 12 times the number of years plus the number of months, yielding the three variables. Where this cannot be computed (both years and months blanks)leave blank for modal imputation by the type III program.

The type III program should also generate additional variables as follows:

1. A supervisor relevance variable equal to 1/12 the sum of the three variables from items 4-6 (the sum rounded to nearest integer) plus the coded values from items 3, 7-9, and the 1-4 ones in the dichotomies from item 10. This computation should follow the modal imputations and precede step 2.

2. Multiply each variable score on those variables from items 11 through 19 by the supervisor relevance score developed in step 1.

3. Output on interim file 3, both the original scores and those weighted by "supervisor relevance".

Range checking operations in program II should be applied in processing both MEII and MJEF. Range checking should also be applied in the type III program for MJEF on the generated variables.

Table 5

Data Entry Specifications for MJEF

Item No.	No. of Variables	Codes	Blanks	Multiple Responses
3	1	2-0	Mode	BM
4-6*	3	No. of months	Mode	NA
7	1	0-5	Mode	NA
8	1	3-0	Mode	NA
9	1	5-1	Mode	NA
10	4	1/0	0	NA
11	1	3-1	Mode	BM
12	1	5-1	Mode	BM
13	1	5-1	Mode	BM
14	1	3-1	Mode	BM
15	1	3-1	Mode	BM
16	5	2-0	1	1
17	6	1/0	0	NA
	+GenVar	0-4 for B-F	Mode	BM
18	1	0-2	0	1
19	31x3	2-0	1	1

Battery Merge Problems for MEII and MJEF: Respondents and Nonrespondents

Nonrespondents to MEII will have no data for either MEII or MJEF. Thus, they will not be on either interim file 3 by IFID numbers. Prior to merging with the student master file, merge these two interim files together, leaving blank those fields from the supervisor file for students responding to MEII but failing to give a supervisor name, or having done so, the supervisor failed to respond to MJEF. Replace the IFID numbers with PID numbers for merge with the student master file.

When merging with the student master file, retain the blank fields for those with missing data either from MEII or MJEF nonresponse, but program to add a dichotomous variable, 1 if student responded with a usable MEII and zero if not, and another dichotomy, 1 if supervisor responded with a usable MJEF, zero if not (regardless of reason why not). These codes will be needed as criterion variables in developing the regressions required to develop weights for offsetting nonresponse bias in analysis. This operation must be external and will be described in another document on sampling and weighting. It must be repeated for each followup cycle.

A Final Suggestion for Reducing the Load in the Impact Space

In addition to the earlier suggestion to dispense with the cross-sectional impact operations, it is suggested that the 3-year and 5-year followups be combined into a single 4-year followup with minor editorial changes in the nearly identical Forms B of MEII. Because of the tremendous length of MEII, some priorities should be established about what to retain or omit to shorten this monstrous instrument. Some of the more intrusive items might have lower priority if not critical to economic analysis, and the military portion shortened to one or two items. The number of books and magazines not related to job might have lower priority. Part-time employment information sought might be reduced. Retain as high priority the economic amounts, except for the contribution to net worth which can be computed programmatically as indicated in the specifications.

These suggestions will not only reduce the load on the respondent but also the processing load for MISOE and possibly increase response rates.

VIII. Data Entry Specifications for the Teacher and Administrator Batteries

The great similarity in instrumentation for the teacher and administrator batteries makes it convenient to specify the data entry operations for these two batteries in a single chapter. Both batteries start with a cover sheet, at least for initial administration; both are readministered annually, except for the IQ test. The latter instrument has not as yet been chosen and therefore no further specifications can be given at this time, except that it will be administered only on first contact and assumed to be processed in the usual manner with a table lookup operation in the type II program to convert raw scores to an IQ metric.

It is anticipated that the cover sheet will be treated as usual, generating a teacher and an administrator name and address file and a scramble file in each case, with the usual link agency involvement, all in accordance with general specifications given in Chapter II. Similarly, both batteries will have master identification forms (TMIF and AMIF, which should be designated as TMIFU and AMIFU in the replication batteries, respectively).

All forms are assumed to be administrable on or with optically scannable answer sheets to be processed under quality controls specified in previous chapters. It remains, then to discuss the processing of interim files 1 for each instrument. The general specifications for layouts and treatments of gains and losses over time for the teacher and administrator files (both interim and master) were given in Chapter II. Administrative directions for both batteries should be prepared and considered part of the battery logistics.

Specifications for Processing Answer Sheets for the Cover Sheet, TMIF, and AMIF

Processing of the cover sheets and master identification form information should follow the principles delineated at the beginning of Chapter IV for the student input battery. The same attention to detail and quality control operations apply to the Teacher and Administrator battery processing.

The Planning Activities Sheet

If this instrument is to be administered with a scannable answer sheet, two problems must be solved. First, the number of days absent for the reporting week should not be a write-in, but a 0-5 (or 0-6) coded multiple choice marking item. Second, more serious, a write-in of "other" activities would seem possible only with some complex alphabetic gridding. Either we must convert this item 10 into objective, scannable format on the basis of pretest information, or plan to administer the instrument as such without an answer sheet for coding and verified keypunching, with preliminary inspection of returns to decide how to code the open-ended responses. Except for this, the thirteen line items, each generate three variables containing two digits of hours recorded. Impute 00 for missing information, including those unused subfields in item 10. Only range checking and inspection distributions are required. The instrument appears in both batteries.

The Image of Vocational Education

This instrument, in both batteries, lends itself readily to scannable format. If items are placed on the answer sheet, there are several instances where space can be saved by deleting from the stems "I believe" or "In my opinion" and retaining the substantive part of the statement for which agreement or disagreement is indicated. The 28 items yield 28 direct scores which should then be summed to yield a generated total score indicating positive degree intensity of attitude toward vocational education. Because the statements are worded sometimes positively, sometimes negatively with respect to a favorable attitude, to break response set, positive items should be coded 5-1 for SA to SD and negative items reverse coded 1-5 for SA to SD with "3" imputed a priori in all items for missing or multiple responses. The positive items are: 2,4,6,8,9,11,16,18,20,22,24,25,27, and 28. The negative items are: 1,3, 5,7,10,12-15,17,19,21,23, and 26.

The Teacher Program Questionnaire

This instrument is in the teacher battery, but not in the administrator battery. It is identical, except for minor edits in directions, to the Student Program Questionnaire and can be processed on the same specifications: 20 semantic differential scales coded 1-7 or 7-1 depending on the left-right orientation of positive-negative attitudes toward the program, with "3" imputed a priori for missing or multiple response.

The Purdue Teacher Opinionaire

Given only to the teachers as a measure of morale, this instrument purports to yield 10 factor scores and a total score. In the absence of a manual or scoring keys, an attempt was made to provide an a priori coding scheme. This was not successful, or rather was still ambiguous both with respect to factor contribution (under the assumption of independent keying) and whether items were to be coded 1-4 or 4-1. The latter is fairly unambiguous given the positive or negative relation to a general positive morale score. However, several of the factor component names have a negative orientation, so that it is not clear whether the items are to be keyed +/- with respect to factors and factors keyed +/- with respect to total score or not. Moreover, it is likely that the official keys are empirical. Therefore, precise coding specifications will not be given here, but should be developed with actual scoring keys available. Adaptation of the instrument for scanning appears very feasible. Generation of factor and total scores may have to be done in the type III program after imputation of modal codes for missing and multiple response data (despite instructions to answer all items).

The MOETS and MAI

The Massachusetts Occupational Education Teacher Survey consolidates content from the previously planned Massachusetts Teacher Inventory and the Occupational Education Questionnaire. It will be administered to teachers. There are two sub-parts labeled, "Initial Data" and "Follow-up Data", respectively, with item numbering from "1" within each part. Although there seems to be no reason why both parts could not be administered both on initial contact and on replication contacts, this appears to be two forms of the MOETS. In any case, the instrument appears readily adaptable to scanning operations.

The Massachusetts Administrator Inventory is a very brief instrument with all but two of the items having parallels in the MOETS. Its specifications will therefore be given in the same table with those for the MOETS. Again, the MAI is entirely adaptable to scanning.

The current forms for both MOETS and MAI have no items for age, sex, race, or marital status. It is assumed that these items have been shifted to the TMIF and AMIF and can be processed in the same manner as similar items in the student battery. Note, however, that in the age item, at least, additional higher age categories may be involved for teachers and administrators and therefore the coding range has to be extended accordingly.

The coding and editing specifications for these instruments are summarized in Table 6. On two of the items the smaller coded value is recommended in case of multiple response. Item I-11 (I for the initial data part of MOETS) has seven categories counting the last or "none" category; its counterpart, item 6 in MAI, does not have this. In view of the dichotomous coding, it is not necessary and should be deleted from MOETS. If retained, it should be added to MAI and the number of dichotomies becomes 7 instead of 6.

Table 6

Data Entry Specifications for MOETS and MAI

MOETS Item No.	MAI Item No.	No. of Variables	Codes	Blanks	Multiple Response
I-1	1	1	0-5	0	BM
I-2	5	5	0-4	0	SV
I-3	7	1	1-4	Mode	BM
I-4	8	1	1-4	Mode	BM
I-5	9	1	1-4	Mode	BM
I-6	10	1	1-4	Mode	BM
--	3	2x7	0-6	0	BM
--	4	1	0-6	0	BM
I-7	--	2	00-10	0	SV
I-8	--	1	0-3	0	BM
I-9	--	1	0-7	0	BM
I-10	--	1	0-9	0	BM
I-11*	6	6	1/0	0	NA
F-1	--	4	1/0	0	NA
F-2	--	3	1/0	0	NA
F-3	--	2	1-3	0	2
F-4	--	2	00-13	Mode	BM
F-5	--	3	1-5	Mode	BM
F-6	--	16	1-5	0	BM
F-7	--	2	0-5	Mode	BM
F-8	--	1	0-3	Mode	BM
F-9	--	6	0-4	0	BM
F-10	--	1	0-5	0	BM
F-11	--	1	0-5	0	BM
F-12	--	2	0-3	0	BM
F-13	--	4	1/0	0	NA
F-14	--	4	1/0	0	NA
F-15	--	1	2-0	0	1
F-16	--	2	1/0	0	NA
F-17	--	1	2-0	0	1
F-18	--	1	0-4	0	BM
F-19	--	2	0-4	0	BM
F-20	--	1	2-0	0	1
F-21	2	2	1-8	Mode	BM
F-22	--	3	1/0	0	NA